DOCUMENT RESUME

ED 100 798 95 SP 008 683

AUTHOR Wang, Margaret C., Ed.
TITLE The Use of Direct Observation to Study Instructional-Learning Behaviors in School Settings.
INSTITUTION Ford Foundation, New York, N.Y.; Pittsburgh Univ., Pa. Learning Research and Development Center.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE 74
NOTE 168p.; Proceedings of a Symposium of the American Educational Research Association (New Orleans, Louisiana, 1973)

EDRS PRICE MF-$0.75 HC-$7.80 PLUS POSTAGE
DESCRIPTORS *Behavior; *Classroom Observation Techniques; *Learning; *Learning Processes; Symposia

ABSTRACT

This monograph comprises five papers presented at a symposium whose primary purpose was to discuss the role of direct observational techniques in the study of certain learning outcomes and of certain learning and instruction process variables in school settings. Each study represents an attempt to apply direct observational techniques to the investigation of the instructional-learning processes and to use the observational data for evaluating and documenting a wide range of learning outcomes. The first two papers, each examining a different technique, focus on how observational techniques are used as an evaluative instrument to study classroom behaviors across diffuse geographical locations and/or across different educational programs. In the three succeeding papers, the authors report studies designed to focus on certain instructional-learning behaviors for a particular purpose other than for the evaluation of classroom behavior in a summative sense. The last section contains two discussions that offer critical comments on the studies reported and suggest implications for further research. A "Description of the Learning Research and Development Center's Individualized Instructional Programs"; definitions; and a description of the observational instrument and techniques, data recording forms, and the criteria for their use are appended. (PD)

THE USE OF DIRECT OBSERVATION TO STUDY INSTRUCTIONAL-LEARNING BEHAVIORS IN SCHOOL SETTINGS

Margaret C. Wang

Editor

Learning Research and Development Center

University of Pittsburgh

1974

The papers and discussions presented herein constitute the proceedings of a symposium held at the 1973 meetings of the American Educational Research Association, New Orleans, Louisiana. Preparation of the manuscript was supported by grants from the Ford Foundation and by the Learning Research and Development Center supported in part as a research and development center by funds from the National Institute of Education, United States Department of Health, Education, and Welfare. The opinions expressed do not necessarily reflect the position or policy of the sponsoring agencies and no official endorsement should be inferred.

TABLE OF CONTENTS

FOREWORD

Teaching, like other professional endeavors, rests upon a foundation of accumulated knowledge and artistry of the practitioner. Any analysis of the teaching process and the activities of teachers must keep both knowledge and artistry in mind. Attempts to study classroom practices must probe into both these levels: general principles that underlie the ways in which learning can occur, and nuances of technique--imposed by teacher or student behavior--that either enhance or deteriorate the instructional model being implemented. In addition, studies of those unexpected emergent processes that result when particular practices are carried out in a school setting are also in order.

It is apparent that process variables are the key to understanding what goes on in schools. In the recent past, educational research studies frequently reported that using method A in the classroom was more effective than using method B, or that a comparison of method A and a control group showed no difference in their effects. Studies of this kind carefully described (1) the characteristics of the student population; (2) the kind of teaching variable manipulated (e.g., instruction with or without TV, individualized versus conventional instruction, or a permissive teacher as compared to a more directive teacher); and (3) student outcomes (e.g., achievement or attitude). Conclusions were then drawn about the influence of the intervening teaching variable. Upon reading these accounts, the nagging question that always came to mind was: What went on in the classroom? (What were the processes that differentiated TV teaching from teaching without TV, or that distinguished permissiveness from directive instruction?) Most studies described the practices that were devised for classroom implementation. But, the details of how these practices were carried out were rarely so carefully analyzed as student

differences or educational outcomes for which tests were either available or specially developed. Differences in the degree of implementation of the teaching variable that required the design of detailed observational procedures were rarely specified and assessed in any precise fashion. As a consequence of these studies, some practical decisions were made for or against the teaching practice involved, but little was learned about the underlying processes that intervened between student characteristics and student outcomes.

The merit of this symposium is its emphasis on the description of process variables. In their papers, all authors realize that improvements in teaching require cumulative knowledge based upon the careful observation of instructional-learning processes in school settings. To be a useful contribution to cumulative understanding, this knowledge must have certain characteristics. First, it must be articulated in terms of generalized conceptual categories that provide a framework for focusing observation and analysis. In this way, studies of teaching can develop richer interconnected theoretical structures than have come out of past studies carried out with few conceptions that are linked to one another.

A second characteristic of descriptions of school processes is that they must be treated in context, i.e., related to student characteristics, teacher styles, school settings, and different educational goals. While teaching processes are widely applicable to the extent that they reflect general principles, their application demands that the principles be adjusted to the conditions of specific situations. The application of a general principle or scientific law must always include parameters relevant to the local setting. In their studies of the processes of teaching and learning, the authors of the papers in this symposium are sensitive both to the requirements of generalizability for knowledge accumulation and of specificity for describing actual practice.

There are several ways to read this symposium. My own recommendation is that the reader start at the end, with the discussants' comments and observations, and then sample the papers according to his inclinations. In this way, the reader can get a feeling of the study of process from the point of view of his interests, whether they focus on the design of an observational system, observation as a tool for evaluating the implementation of an instructional model, or observation as a way of studying emergent social structure in the classroom.

The papers in this symposium carry us another step toward a future in which teaching and evaluation research in education will no longer be isolated collections of classroom studies but will instead be investigations of classroom processes based upon systematic conceptually based observational schemes. As a result, groups of studies will relate to one another, and a coherent and improved understanding of the processes at work in classrooms will emerge.

Robert Glaser
Pittsburgh, Pennsylvania
November, 1973

INTRODUCTION

It is widely acknowledged that because of the lack of adequate techniques, a great number of important variables related to schooling--instructional and learning processes, as well as learning outcomes--are often ignored and seldom systematically studied. In spite of the fact that in recent years, educators and educational researchers have been stressing the need to develop techniques that provide information for those variables for which the traditional paper and pencil techniques are neither sufficient nor appropriate; this traditional practice still prevails. The papers included in this monograph are examples of research studies designed specifically to meet the urgent need for the development of alternative techniques in the investigation of variables related to what goes on in schools.

The papers are proceedings of a symposium, presented at the Annual Meeting of the American Educational Research Association in New Orleans, Louisiana, February 1973. The primary purpose of the symposium was to discuss the role of direct observational techniques in the study of certain learning outcomes, and of certain learning and instructional process variables in school settings. Each study represents a unique attempt to apply direct observational techniques to the investigation of the instructional-learning processes, and to use the observational data for evaluating and documenting a wide range of learning outcomes.

The first two papers, each examining a different technique, focus on how observational techniques are used as an evaluative instrument to study classroom behaviors across diffuse geographical locations and/or across different educational programs. Both techniques, originally designed to assess the degree of implementation of a particular educational program, were used in Project Follow Through, a nationwide

program initiated by the United States Office of Education in 1967. The program was initiated as an effort to improve the education for low income children. Through the development of innovative educational programs for the first four grades in the American public schools (kindergarten through grade three), Project Follow Through attempts to bring about educational reforms in the public schools. The Bank Street College Follow Through Program, described in the Ross and Zimiles paper, and the Learning Research and Development Center (LRDC) Follow Through Program, described in the Leinhardt paper, were among the first programs developed under the federally-sponsored project.

The Differentiated Child Behavior Observation System (DCB), described in the paper by Ross and Zimiles, was developed to obtain descriptive information about classroom interactions in order to compare and differentiate program impacts on students' classroom behavior (cognitive, affective, physical). Leinhardt, in "Observation as a Tool for Evaluation and Implementation," deals specifically with the use of direct observation as a means to monitor the implementation of a particular program. Leinhardt uses the observational data to measure and compare differences and similarities among classrooms using the same instructional program.

In the three succeeding papers, the authors report studies designed to focus on certain instructional-learning behaviors for a particular purpose, other than for the evaluation of classroom behavior in a summative sense. Wang reports on a study designed to illustrate the use of direct observational techniques in formative evaluation of an instructional-learning management system developed to implement an individualized instructional program. Wang specifically emphasizes the application of observational data to the development and refinement of the instructional

program. Shimron reports on a study designed to investigate the effects of an educational program on individual students for whom certain student learning characteristics are known. Shimron focuses on the use of observational data to document behavior patterns of the individual student in classroom learning situations. The papers by Leinhardt, Wang, and Shimron are related in the sense that they are concerned with the evaluation of certain aspects of the same educational program, the LRDC Instructional Program. For a brief description of the LRDC program, see Appendix A.

Omark and Edelman discuss the application of ethological concepts and methods to study the social behavior of young children. The authors suggest a method, which is well-established in the naturalistic study of the social behavior of primates, as one way to obtain a broader and deeper study of children's concepts of their social world, and their interrelationship with peers.

The last section includes remarks by Medley and Resnick who were discussants for the symposium. The remarks offer critical comments on the studies reported, and suggest implications for further research in this area.

It is hoped that this monograph, reporting examples of fruitful use of direct observational data for studying the instructional-learning processes and outcomes, will demonstrate the possibility of and the need for broadening the base of school evaluation, as well as the development of a variety of alternative techniques to investigate and evaluate certain aspects of schooling.

Margaret C. Wang
Pittsburgh, Pennsylvania
November, 1973

THE DIFFERENTIATED CHILD BEHAVIOR OBSERVATIONAL SYSTEM[1]

Sylvia Ross and Herbert Zimiles

Bank Street College of Education

In recent years, researchers have increasingly sought to characterize classroom interactions in a systematic and relatively objective manner. These efforts can be justified in terms of the need for a description of the life experience of the child in the classroom. Although it is recognized that the classroom constitutes a significant influence on the child's development and is a major source of each individual's introduction and exposure to society, we lack even basic information about the nature and content of classroom interactions. However, a number of critical issues have generated these efforts: (1) the search for generalizable information about teacher effectiveness and classroom climate that could be used both for teacher training and evaluation purposes, and (2) the growing discontent with using applied standardized tests for evaluating educational programs. The increasing recognition of the limitations of these measures, particularly in reference to assessment of the effects

---

[1] The research reported in the present form is supported by a grant from the Ford Foundation to the authors. The authors would like to acknowledge Elizabeth Gilkeson who made substantive contributions during the development of the DCB; Barbara Biber, Margery Franklin, and Carda Bowman for the valuable suggestions and critical comments regarding particular aspects of the observation system; and Dinah Heller and Michael Moss for their assistance in supervising the data gathering by a team of graduate students.

of federally-funded programs on inner-city children's cognitive abilities, has stimulated a search for broader and more encompassing evaluation and program analysis measures.

Overview of the System

The Differentiated Child Behavior Observational System (DCB) was originally developed under the auspices of a Follow Through sponsor--Bank Street College of Education--as one of a number of self-evaluation measures designed to assess the extent to which its own program had been successfully implemented. The Bank Street sponsorship of Follow Through classes in 14 communities in the United States involves the implementation in inner-city public school classrooms of a "developmental-interaction," open-classroom approach (see Shapiro & Biber, 1972) developed and applied over many decades in the College's School for Children.

The complexity of life in the informal or open classroom makes recording of all relevant information difficult. The term "relevant" used in this context defines the extent to which choices are made on the basis of values, or of judgments as to which behaviors will provide criterion measures of children's classroom behavior. The content of the DCB instruments reflects a set of assumptions and values underlying the Bank Street approach (Gilkeson, 1970), with "competence . . . conceived functionally in terms of how the individual interacts with the challenges, the people and the work of his environment" (Biber, Shapiro, & Wickens, 1971). It also reflects an attempt to delineate a comprehensive and detailed roster of typical classroom interactions. The basic assumption underlying the design of the DCB is that children's behavior (cognitive, affective, social, and physical) reflects the attitudes, values, and curriculum foci of the classroom instructional team.

The Open Classroom versus the Traditional Classroom

How are the differences between an open and traditional classroom reflected in children's interactional behavior? What is the effect of informal spatial arrangements and greater teacher and pupil mobility on the quantity and quality of classroom interactions? Does an "open," independence-fostering, child-centered environment, that supposedly encourages self-expression, produce a greater incidence of destructive, acting-out behavior than a setting which has a high degree of control as one of its major practices? Does the attempt to integrate and balance cognitive, affective, aesthetic, and social learning experiences result in less cognitive involvement than that found in traditional settings where academic learning is the primary objective?

The Relation of SES and Ethnicity to Program Patterns

To what extent does the SES and ethnicity of a school population affect quantity and quality of classroom interactions? How does the behavior of the inner-city child in the open classroom differ from his behavior in a traditional setting? How does the behavior of the low income child differ from that of the middle-class child in either setting?

The Observational System

The DCB is used "live" in the classroom. Data are gathered by trained observers who encode children's interactional behaviors on timed and change-of-behavior bases. One of the distinctive features of this system is its emphasis on the substantive aspects of children's interactions: it provides data regarding the content as well as the source and direction of each entry. In addition, unlike many previously designed observational approaches, it incorporates a number of procedures that

are applicable both to informal, open classes as well as to more traditional settings. The observational system includes two instruments: the DCB Form and the Classroom Scan. (See Appendix B.)

The DCB Form has been designed to provide quantitative and qualitative data regarding children's verbal and nonverbal classroom behaviors. The focus (whether in observations of small, large, or total groups) is on the number of occurrences of specified behaviors as well as on the nature of the interaction in each instance, i.e., child-to-child, child-to-adult, to or by self; adult-elicited or child-initiated; individual or choral response. The referent child's sex is also indicated in each instance.

The six major behavioral categories of the DCB are: Gives Information, Asks Questions, Expresses, Acts Destructively, Organizes and Manages, and Represents and Symbolizes. Each of these six categories includes from six to nine subcategories that are designed to identify specific behaviors within each general category.[2]

The first two categories, "Gives Information" and "Asks Questions," are primarily concerned with verbal behaviors in the cognitive domain. The subcategories subsumed under these headings have, to some extent, been ordered according to their complexity.

Category 1: "Gives Information"

Subcategory 1: Identity-Situation includes factual information regarding personal events ("I got a puppy for my birthday") and labeling without further descriptive or differentiating details.

---

[2] This presentation is limited to the most salient points of definition. A comprehensive listing of examples for each subcategory component is presented in the Observer's Manual.

Subcategory 2: Prediction, Plan includes the child's guess or hunch ("I'll bet there's a frog in that jar") and projections of future plans ("Tomorrow I'm bringing all the stones I collected and I'm going to start labeling them for our science table").

Subcategory 3: Function, Process, Instructions includes information regarding the functions of certain items ("The thermometer is to see how hot the water is"), how things work ("When you strike the key, it pushes the lever"), and how to carry out a task or play a game ("The double checker can move either way").

Subcategory 4: Differentiating Properties covers a wide range of descriptive statements regarding sensory qualities such as color and texture and defining characteristics, e.g., size, form, or quantity ("There are 10 fish in the tank").

Subcategory 5: Relationships deals with comparisons ("It's the same color as the moss") as well as with temporal, spatial, or ordinal relations ("There are 60 seconds in a minute").

Subcategory 6: Category, Class includes behaviors in which group membership is identified ("Mold is a kind of plant").

Subcategory 7: Causal Reasoning, Problem Solving includes an attempt to explain why things happen the way they do ("It's heavier than water; that's why it sinks"), and a solution to an identified problem ("What we need are some railings so the cars won't go off").

Category II: "Asks Questions"

The subcategories in Category II are the interrogative parallels of those described above for Category I. For example, if the child points to an object asking "What's that?" it would be entered in Category II,

Subcategory I, Identity-Situation; the question "Which is bigger?" would be entered in Subcategory 5, Relationships. Subcategory 7, Causal Reasoning, Problem Solving includes the child's inquiries regarding the underlying cause of natural phenomena ("What makes the lightning?") or an observed event ("Why do some things stay on top of the water and not others?").

Category III: "Expresses"

Category III includes both verbal and nonverbal behaviors that are primarily affective in content, i.e., expressions of feelings and attitudes (negative as well as positive) and of preferences and needs.

Subcategory 1: Routine Needs refers to behaviors such as pencil sharpening or taking a drink of water.

Subcategory 2: Needs - Social/Physical includes complaints ("Your paper is taking too much room"; "My knee hurts where I bumped it"), and requests ("Move over so I can see").

Subcategory 3: Needs - Task Related includes general requests for assistance, materials, or equipment ("We need stuff for a collage"), and requests for approval or recognition of one's own work.

Subcategory 4: Preferences and Desires includes responses given within the context of explicit choice as well as more general expressions of individual preferences ("I hope we have tuna for lunch").

Subcategory 5: Feelings, Attitudes, Opinions includes negative as well as positive expression ("I hate cleaning up"), and beliefs ("Girls are smarter than boys").

Subcategory 6: Interest in Another's Attitudes or Opinions includes questions such as "What's your favorite program?" and "Do you like to go to gym?"

Subcategory 7: Affection, Warmth, Humor includes nonverbal behaviors (child strokes rabbit gently) as well as verbal statements indicating a positive relationship with another ("We're best friends, right?"), good-natured joint laughter, and sharing "joke" with pleasure evidenced by all concerned.

Subcategory 8: Concern for Others includes nonverbal behaviors showing willingness to help or share as well as verbal statements or questions indicating support of another's efforts, or consideration for another's feelings and well-being ("You can have some of mine").

Subcategory 9: Unwillingness to Help or Share includes negative responses to requests for help ("Go get it yourself") or for sharing belongings, materials, or equipment ("Don't give them any; they're just for us").

Category IV: "Behaves Destructively"

Category IV is rigorously defined to refer to behaviors in which there is direct evidence of physical abuse or threat of physical abuse.

Subcategory 1: Verbal - Initiates includes threats of physical abuse as well as extreme taunting, derisive behavior ("You're the stupidest one in the class; you never know any of the answers, dumb-dumb").

Subcategory 2: Verbal Defense refers to retaliatory threats ("If you do that, I'll punch you right back").

Subcategory 3: Physical - Initiates is limited to those overt behaviors that appear intended to cause injury.

Subcategory 4: Physical Defense is limited to retaliatory behaviors that appear intended to cause injury.

Subcategory 5: Takes Other's Belongings is limited to overt behaviors that appear intended to be destructive.

Subcategory 6: Destroys Another's Work, same as Subcategory 5, above.

Subcategory 7: Abuses Material and Equipment, same as Subcategory 5, above.

Subcategory 8: Challenges Established Classroom Rules includes negative responses by child to reminders regarding established classroom limits and expectations.

Category V: "Organizes and Manages"

Category V represents the attempt to assess the extent of child autonomy evidenced in the classroom.

Subcategory 1: Records Choice of Task (on bulletin or blackboard).

Subcategory 2: Suggests Task or Activity ("Let's play with the blocks").

Subcategory 3: Initiates Task (Child goes to an easel and starts to paint).

Subcategory 4: Commands, Directs refers to managerial rather than instructional behaviors ("Get the blocks").

Subcategory 5: Initiates Attentional Focus refers to a situation where the child enthusiastically calls attention to an event or feature that he believes is of general interest ("Hey, look at the plant; there's another green thing coming out"). It should be noted that calling attention to one's own work would be entered in Category III, Subcategory 3: Needs - Task Related.

Subcategory 6: Seeks Answers requires some evidence that a question has been raised and a possible resource identified ("Let's look at the chart--it tells you how much you need").

Subcategory 7: Selects Materials refers to the selection of materials and equipment from the cabinet or shelf where they are stored.

Subcategory 8: Replaces Materials, the child returns material and equipment, as above.

Subcategory 9: Straightens Up Work Area includes behaviors such as mopping up spills or picking up puzzle pieces that have fallen onto the floor.

Subcategory 10: Attempts to Resolve Conflict refers to an attempt by the child to reduce conflict by clarifying a situation ("We didn't mean to knock it over; it was an accident") or an attempt to settle a dispute by compromise ("Since they won't let us make it longer over there, we'll make it longer over here").

Category VI: "Represents and Symbolizes"

Category VI focuses on the child's aesthetic and imaginative expressions as well as symbolic interactions.

Subcategory 1: Structures Dramatic Episode includes behaviors relating to dramatic play in which the child may assign roles and give directions ("Let's pretend this is the bus and I'll be the driver").

Subcategory 2: Elaborates Dramatic Episode includes role-playing and use of objects to represent other things.

Subcategory 3: Makes Descriptive Comparisons includes the use of figures of speech or idiosyncratic modes of description ("I feel cold like a butterfly"--shivering child).

Subcategory 4: Improvises Sound refers to play with syllables, initial consonants, words, or tapped out rhythms, etc.

Subcategory 5: Makes up Story, Song, Poem includes child's creative expressions whether in the form of story, song, poem, or dance.

Subcategory 6: Tells Familiar Story may also include the sharing of a riddle or familiar song, or part of a TV show.

Subcategory 7: Narrates Sequence of Events refers to both personal and historical events.

Subcategory 8: Shares Symbolic Experience refers to sharing pleasure in a book or picture with emphasis on the content.

Subcategory 9: Decodes/Pronounces (for practice) is limited to a more routine type of reading out loud to another as part of skill practice with emphasis on decoding practice.

A single DCB Form is used for each five-minute interval of observational recording with a total of 12 DCB Forms used for one day of observation. The observations follow a systematic course that is designed to provide representative samples of the behaviors of all the children in the classroom as they are observed in groups of various sizes and participating in ongoing activities with and without adult intervention. Activity, grouping, and adult role are indicated for each DCB Form.

At the end of each five-minute observation period, the observer completes a brief rating scale indicating the extent to which coded entries were activity-related and the degree of task persistence encountered.

The Classroom Scan provides a measure of the behavior of each child in the classroom during each of six time samples during the day (i.e., whether involved in an activity, observing, involved in a social-physical interaction, destructive act, or showing "no observable focus").

It also provides for a description of the number and kinds of all ongoing activities and groupings, identifying adult role, if any (i.e., supporting or directing), in relation to each group. The activities are further described as to the perceptual modes involved, degree of abstraction, and dimensionality. Each scan is administered immediately after a set of two DCB Forms has been completed with a total of six scans for each day of observation.

Overview of Previous Findings

Preliminary work with the DCB was based on three groups of elementary school children age five through eight. The first group was drawn from the Bank Street School for Children, which has an open-classroom approach, and consisted primarily of middle-class children. The second group came from the Bank Street Follow Through classes in inner-city public schools--also with an open-classroom approach. The third group of children attended non-Follow Through inner-city public schools with a traditional classroom approach. Thus, three distinct reference points were provided for examining the DCB data.

Results

The results are summarized in Table 1. Substantial differences were found among the three groups from which data were drawn. The Bank Street School for Children classrooms totaled twice the number of interactions found in the traditional public school classes, the number of interactions in the Follow Through group was closer to the number in Bank Street School rather than in the non-Follow Through group, showing 60 percent more entries than the traditional classes ($p < .01$). The total number of children's interactions recorded in each setting provided one means to describe a classroom environment along a passive-active

Table 1

Mean Category Scores by Group

Category	Bank Street School for Children Classrooms N = 2	Bank Street Sponsored F/T Public School Classrooms N = 40	Traditional Inner-City Public School Classrooms N = 6
I. Gives Information*	148.4	122.8	52.4
II. Asks Questions*	30.1	24.1	5.1
III. Expresses	109.4	109.3	66.1
IV. Behaves Agressively	1.9	10.9	10.0
V. Shows Autonomy	43.5	20.1	7.0
VI. Represents and Symbolizes**	55.7	38.5	28.7
TOTAL***	428.7	344.7	215.7

* Subcategories 2 through 7 with Identity Situation omitted

** Adult-elicited behaviors omitted

*** Because of the omissions cited in the above footnotes, the total scores do not constitute a sum of the frequency scores for the six categories.

continuum. In these results, the Bank Street School for Children represents one polarity--"very active"--and the traditional public school classes represent another--"least active" or "passive"--with the Follow Through classes notably more active than passive; the traditional public school classes were more passive.

It should be emphasized that although this analysis is limited to a measure of interactional behaviors, there is no assumption that noninteractional behaviors, e.g., the child reading or writing by himself, are either passive or nonproductive. However, since observers sought to record children's behaviors in those situations that fostered interaction, the results would appear to reflect the general level of interaction possible within a given classroom. We may conclude from the sharp differences in the total scores that the children in the traditional classes were either far less active, or were engaged in tasks that provided less opportunity for interactions.

In all three types of classrooms; the most frequent interaction was that described as giving information, i.e., Category I: Cognitive Domain. The traditional public school classes showed by far the fewest of the higher-order cognitive interactions. The Bank Street Follow Through Program had more than twice the number of higher-order cognitive interactions ($p < .01$) and the Bank Street School for Children had almost three times as many ($p < .001$).

The second cognitive category, that concerned with asking questions, occurred much less frequently. When examined on a proportional basis, differences among school groups were even more massive than those found with regard to giving information. The children in the Bank Street Follow Through classes asked more than four times as many questions as the children in the traditional public school classes, and the Bank

Street School children were found to ask questions more than five times as often as those in Follow Through and the traditional public school ($p < .01$, $p < .001$, respectively).

The second most frequent category of behavior recorded by our system was that concerned with the expressiveness of children (Category III). The Bank Street Follow Through groups and the Bank Street School classes showed approximately equal amounts of expressive behavior, but both frequencies were almost twice that found in the traditional public school classes ($p < .04$, $p < .01$, respectively). When examined on a proportional basis, differences among the three school groups in expressive interactions were relatively small, especially when compared to the large differences found among the school groups in cognitive interactions.

Another interactive category that yielded large differences among the school groups was that concerned with autonomous behavior (Category V). By far the least amount of autonomous behavior was shown by the traditional public school groups. Those enrolled in the Bank Street Follow Through Program showed almost three times as many such behaviors, while the Bank Street School children showed more than twice as many autonomous behaviors than the Bank Street Follow Through children and more than six times as many autonomous behaviors than the traditional public school children ($p < .01$).

Among our most surprising findings was the low incidence of destructive behavior found in all three groups.[3] The largest difference was between the Bank Street School and the two public school groups ($p < .01$);

---

[3] The incidence of such behaviors is reflected by the following mean total frequencies for the three groups: Bank Street School = 1.9, Follow Through = 10, and Comparison = 10.

there were five times the number of destructive interactions in both the Bank Street Follow Through classes and in the traditional public school classes. However, when these findings are considered in relation to the proportion of total behaviors represented by each group's score, the number of destructive acts in Follow Through actually represents a considerably smaller proportion of its total behaviors than does the traditional group's similar score.

Although the number of adult-elicited behaviors was similar in the three groups, there was a far greater number of child-initiated behaviors in the Follow Through and Bank Street School groups than in the traditional classes ($p < .001$). However, those in the Bank Street School showed more child-initiated child-to-adult behaviors than did those either in Follow Through or the traditional public school classes.

In sum, our preliminary data indicate that the DCB shows promise of contributing to a body of information that will ultimately describe the ecology of the classroom. In spite of the dramatic differences in classroom management and teaching styles in our sample, a number of uniform patterns of classroom interaction were observed. The most frequent category of classroom interaction, in all cases, was that of giving information. This form of cognitive interaction occurred most often. According to our method of categorizing classroom interaction, the second most frequent form of interaction was that concerned with expressive behavior. In all of the schools we studied, destructive behavior was found to occur relatively infrequently among the children's interactions. The most striking differences found among the groups were in the total amount of interaction observed, the amount of conceptually-based information exchanged, the amount of questioning and autonomous behavior, and the amount of child-initiated behavior.

Although the Follow Through classes were more like the traditional public school classes than the Bank Street School in terms of the SES backgrounds and school settings, the DCB findings indicate that the children's interactions in Bank Street Follow Through classes were more like those in the Bank Street School classes than those of the traditional public school. It should be pointed out that the Follow Through classes did not all manifest identical patterns, but could be identified along a continuum, with some classes showing patterns very similar to those of the Bank Street School and some showing patterns closer to those of the traditional public school.

It should also be noted that the DCB records and describes what the observer focuses upon, and thereby captures the quality of interaction that occurs, but does not report the number of activities that are taking place at the same time. Comprehensive coverage of classroom transactions is provided by supplementary observations using the Classroom Scan, which yields a detailed description of all activity-groupings occurring at six time periods throughout the day.

Current Work

Current work on the DCB is concerned with refining the coding procedures and expanding their coverage; cross-validating the previous findings with a new set of equally varied classrooms and extending the comparative study of classrooms to subcategory descriptions as well; examining the path of change in the data recorded by the DCB during the course of the academic year (patterns of classroom interaction obtained from DCBs in the fall will be compared with those revealed in the spring); determining the degree of relationship between DCB scores and an assessment of teacher behavior (an independent assessment of teaching behavior will be related to classroom interaction data obtained from the DCB); and

assessing the reliability of DCB scores by comparing data from two different observers who observe in the same classroom during the same time period, with data based on observations by the same observer on different days. Although previous work with the DCB indicates that it is sufficiently reliable to differentiate sharply among different types of classrooms, it is essential to estimate the magnitude of error attributable to variation among different observers and to day-to-day variation within the same classroom and the same observer.

The foregoing discussion has described in summary fashion one effort to arrive at a method for observing and recording children's interaction in the classroom. DCB data may serve either as the independent variable or the dependent variable in an educational experiment because they describe phenomena that mediate between the educational stimulus and its internalized impact upon children. The DCB data serve as both an index of the quality of educational intervention experienced by a classroom of children and as an indicator of the kind of influence a particular form of educational intervention is likely to have upon its participants.

References

Biber, B., Shapiro, E., & Wickens, D. Promoting cognitive growth: A developmental-interaction point of view. Washington, D. C.: National Association for the Education of Young Children, 1971.

Gilkeson, E. The Bank Street approach to Follow Through. New York: Bank Street College of Education, 1970. Mimeo.

Shapiro, E., & Biber, B. The education of young children: A developmental-interaction approach. Teachers College Record, 1972, 74(1), 55-80

OBSERVATION AS A TOOL FOR EVALUATION OF IMPLEMENTATION[1]

Gaea Leinhardt

Learning Research and Development Center
University of Pittsburgh

In the United States in the past 15 years there has been a marked effort to improve education at all levels with special emphasis at the preschool and elementary grades. This effort has resulted in a proliferation of new subject matter sequences as well as more global educational alternatives such as the "open classroom." The question now being raised is whether or not the programs have been successful; that is, requests are being made to evaluate the effects of educational innovation. The form which the evaluations take has been largely a function of the nature of the innovations. Innovations which focus on the process of education tend to emphasize classroom descriptions of attitude, climate, and interaction patterns; while innovations which focus on academic improvement tend to emphasize positive changes in standardized subject matter tests.

This paper presents information which can broaden the interpretation and utility of outcome measures on standardized tests. The

---

[1] The research reported herein was supported in part by a grant from the United States Office of Education to the Learning Research and Development Center. The opinions expressed do not necessarily reflect the position or policy of the Office of Education and no official endorsement should be inferred.

purpose of this paper is to demonstrate the need for and means of evaluating implementation of educational innovations. It presents information which can broaden the interpretation and use of outcome measures on standardized tests. Measures of implementation can both clarify the nature of the educational process and demonstrate the relationship of that process to observed achievement. Several assumptions are made. First, educational innovations need to be evaluated not only for the obvious economic reasons, but also in order to provide clearer insight into areas which need improvement. Second, the reporting of educational outcomes without relating the outcomes to the innovative process does not constitute a meaningful evaluation. Third, an appropriate way to evaluate educational innovations is to measure both input or antecedent variables and process or transaction variables, and to use those measures to explain or predict outcomes (Stufflebeam, 1971; Stake, 1967; and Cooley, 1971). The main body of this paper is concerned with indicating what type of information is useful for measuring implementation variables and how to assure that those measures will be credible.

Setting

The Learning Research and Development Center (LRDC) is currently involved in evaluating its educational program in the Follow Through Schools. LRDC is one of 22 sponsors in the nationwide Follow Through program. The Learning Research and Development Center's Instructional Model (see Appendix A) is present in seven Follow Through sites in kindergarten through third grades; each site consists of from two to seven schools. The evaluation effort described here focuses on the second grade classrooms at four established sites, those sites which have had the program at least one year. The input data, which described the entering aptitude of students, consist of the Lorge-Thorndike Cognitive

Abilities Test, a general abilities test. The output data consist of measures on the Wide Range Achievement Test (WRAT). The data for implementation consist of descriptions and measures of the various dimensions of the classroom obtained from an instrument which was specifically designed for that purpose.

All investigations which take place in a natural setting have some unique restrictions and advantages associated with them. The advantages are the tremendously increased credibility and generalizability of the information obtained. Clearly, if one can demonstrate that a program can be implemented and that the implementation improves performance in such widely differing settings as Follow Through, one has built a very strong case for the program. The disadvantages, however, are also very great; they focus on the following three areas: the geographic location of the classrooms (they are widely dispersed nationally); the staffing at each site (it varies in terms of the availability and willingness of its members to engage in evaluation activities); and the record keeping process (no permanent or consistent records of testing and prescription are normally kept).

The Development of an Implementation Instrument

Figure 1 shows the sequence by which an instrument for measuring the implementation of the LRDC Instructional Program was developed and tried out. I view the steps as necessary and sufficient for the development of an implementation instrument; however, I do not view this as a unique solution to the problem of such development. The diagram is read in the traditional manner and will not be discussed in detail, but it will be referred to throughout the paper.

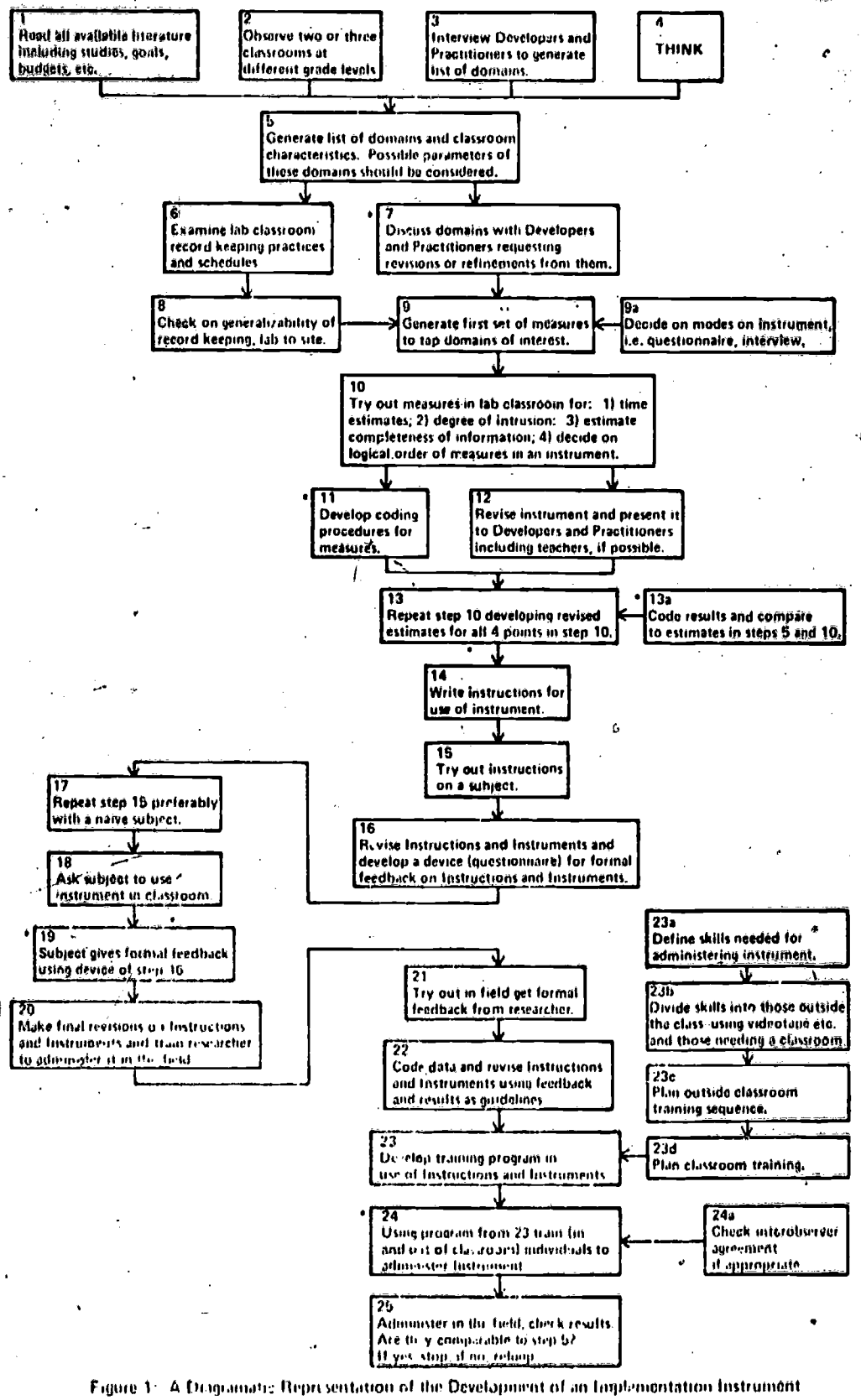


Figure 1: A Diagramatic Representation of the Development of an Implementation Instrument

The first four steps in the figure generated specific information about the LRDC program and about the underlying theories of education on which the model was built. The discussions with developers and implementors brought out the concern that while the measures might be restricted to the unique aspects of LRDC's program, the domains of the variables tapped should be generalizable to a variety of educational settings. Stated somewhat differently, the specific measures are nested within variables (potentially measurable in diverse ways) which in turn are nested within fundamental domains of concern. The following list of the variables that need to be measured emerged from observing classrooms, discussing the program with developers and implementors, and examining the literature: the context variables of each classroom, the allocation of time, the allocation of space, the assignment and measurement procedures, classroom management, and student independence. Measures of these variables should serve the following functions: (1) provide descriptive information about the field sites using the LRDC program; (2) provide a basis for comparing the laboratory and field schools; (3) provide a basis for comparing the model and the field; and (4) provide an explanation of output variance not accounted for by input measures.

The Instrument

After six major revisions of trial instruments, a field reading version was developed. That is, steps 1-15 on Figure 1 were cycled through approximately six times before an instrument which could be used in the field was developed. Following the field test of this instrument, a final version was developed incorporating minor revisions which arose from the feedback of the field tester (Step 19). The instrument itself, the nature of instruction, content validation, and the training program for its use are presented elsewhere (Leinhardt, 1972). The

instrument covers four major areas: background, prescription, testing, and teacher interactions, which will each be discussed briefly.

Background. The first four questions on the instrument provide background information on the classroom; the number of boys and girls involved; the number of children who are presented on the day of observation; the number of years the teacher has been using PEP-IPI; the age range of the class; the size of the class in square feet (transformed into the number of square feet per child), and the allocation of time and space for exploratory. (See Appendix C.)

Prescription. The next area on the instrument concerns prescription information (see Appendix C). Question 5a on the instrument asks the observer to list all of the IPI Math or Quantification assignments written on each child's prescription sheet on the day of the observation. The information is obtained by looking at each child's ticket (sheet) or folder and recording the most recent list of assignments. The question was coded in a manner which would yield information about the uniqueness of the list of assignments obtained. A ratio of unique assignments (different by units and levels) over total assignment was formed, giving a single measure of uniqueness for the classroom.

Testing. The next domain is testing. To gather some information about the procedures being used, question 5b asks the individual to list the dates of the last time a child was tested in any way in math (IPI or PEP). For each date listed, the number of school days from that date to the date of observation is counted, coding the date of observation as zero and skipping weekends. To gain information about the time between testings, the average number of days between testings were obtained from the same prescription sheets used in 5a or from the teacher's testing data in Quantification. Testing is being viewed here as an opportunity for the

teacher to use information about the students' performance to alter the assignments given to them.

The major function of testing is to accurately place and then monitor the progress of the child in the curriculum. From the testing information, prescriptions for the daily activities of the child are written which in turn affect the rate of progress through the curriculum (Glaser, 1968). One of the aspects of testing which is assumed to be important for an individualized program is that it occurs at regular enough intervals to provide accurate information about the changing curricular needs of the child. Unfortunately, regular intervals are not defined anywhere in the literature about implementing the LRDC program. However, from discussing the problem with program developers, there was a consensus that a reasonable time lapse between tests ranged from five to ten days; a one-week lapse was reasonable while a three-week lapse was unreasonable.

Teacher Interactions. In any individualized program, a great deal of the success or failure is dependent upon the teacher's interactions with the students. The teacher is the one who sets the tone or atmosphere of individualization. If his/her actions convey a sense of group rather than individual treatment, then the effort to individualize will have failed in a fundamental way. Question 6 attempts to tap some of the relevant information about teacher and student interactions. The teacher's and aide's interactions were observed during the time of day when prescription work was done. Measures obtained from the sixth question were: (1) the frequency of contact, (2) the content of the contact (i.e., what information is being transmitted from the teacher to the child in terms of intellectual and affective communications), and (3) the distribution of contacts or the way in which the teacher allocates her attention among the students.

Distribution, frequency, and content of the teacher's contacts are measured by observing the teacher for 15 or 20 minutes while (s)he is traveling. During the observation, every separate interaction with a child is coded and placed in one of six cells on a sheet, depending on where the child to whom the interaction was directed was located (see Appendix C). The cells correspond to areas of the classroom; the number of children located in each cell is recorded. From this system of observation, one can obtain measures of frequency and distribution over the class of contacts made. The content of the teacher's statement was coded as: general positive, general negative, management, positive management, negative management, cognitive or subject matter oriented, positive cognitive, negative cognitive, management cognitive, positive or negative management cognitive (Leinhardt, 1972). While these categories do not cover all of the nuances of a teacher's behavior, they are the minimal ones that describe behaviors important to children's learning (Reynolds, Light, & Mueller, 1971).

The coding of the observation section is complex; therefore, the measure will be underlined, followed by a brief description of the coding procedures. Total frequency was coded by counting the total number of contacts made for the entire observation time (15 to 20 minutes), and includes: management, cognitive, checkoff, cognitive-management, unattached positive or negatives, and any uncodeable X's. The frequency for cognitives was obtained by counting all cognitives plus cognitive managements for the entire observation period. The frequency of managements was obtained by counting all other contacts made. Thus, frequency of management plus frequency of cognitives equals total frequency. The percentage of negative contacts was obtained by adding all negatively coded contacts (negatives, negative management, negative cognitive, negative cognitive-management) and dividing that by the total number of

contacts made. (The percentage of positive contacts was so small and unvaried that it was not coded after the field trial was examined.)

There are three distribution measures: total, management, and cognitive. They do not add up to total distribution. The system for coding the total distribution will be explained to provide an example; the other two codes are similar. A distribution was calculated for each cell by the formula $\frac{(O-E)^2}{E}$

Where:

O = The observed total number of contacts for that cell (or observed managements or cognitives).

$E = \left(\frac{T_{co}}{T_{ch}}\right) \bar{Ch}$ = The expected total number of contacts for that cell.

T_{co} = Total contacts made over all cells (or total cognitives, etc.)

T_{ch} = Total number of children in the system obtained by $\sum_{1}^{6} (\bar{Ch})$

$\bar{Ch}$ = The average number of children per cell based on 2 counts or 4 counts depending on the two forms.

The distribution measure is then obtained by summing over all cells. If the observed frequency equalled the expected frequency in each cell, the total score would be zero. Thus, the smaller the measure, the more evenly the teacher and aide are distributing their attention.

The three variables--frequency, content, and distribution--are each important indicators of the teacher's style of interaction. The frequency of contact is a good measure of the "travel" rate in the room. It is also a reasonable indicator of how long children wait before they are able to get the teacher's attention. The major difficulty for observation of the content of a teacher's remarks lies with the decision of which aspects are most relevant. I have chosen to focus on a rather simple distinction between negative statements and all others, and cognitive versus management statements. The collapsing of the existing categories was done for several reasons: first, to increase inter-observer agreement; second, to increase the frequency of the observation of the categories within the limited time of observation, and finally, to focus on the most relevant parts of the teacher's speech. Again, there is the problem of incomplete rather than inappropriate measures. Those parts of a teacher's interaction which would seem most relevant to the student's advancement are those which concern the affective dimension and subject matter content of the communication.

Other Measures Obtained. Several other measures were obtained which have not been included in this general discussion. Some were obtained by directly questioning Educational Specialists[2] after the instrument had been administered, others were obtained from the instrument itself. Additional measures include: the number of adults observed traveling; whether or not the children get their own work (information was obtained by interviewing the child--it provided useful anecdotal information, but was not very generalizable at the classroom level); hours

[2] An Educational Specialist is a person at the Follow Through site who is responsible for implementing and supporting academic and parent involvement programs. There is usually one specialist for every eight or nine teachers.

assigned in math and reading (regardless of whether the reading is an LRDC program) which was obtained from the site specialists; and finally, the number of days the teacher was absent during the year, also obtained from the specialists. These are all measures of implementation which might be relevant to the student learning outcomes.

Measures Omitted. Several of the original measures in the instrument were not used because of a desire to pare down the number of variables being examined to those which seemed to give reliable, useful, and interpretable results. The questions either condensed or omitted concerned exploratory (except those already mentioned), the room map with teacher traveling pattern, the percentage of contacts which were child initiated, and the child's interview. The room map was very useful in interpreting and in some cases correcting the information recorded during the teacher observation.

The one major observational measure which was lost was the percentage of contacts which were child initiated, because good agreement (above 60 percent) was not obtained with observers. A source of the low agreement is the lack of a consistent signaling system across classrooms. In most classes, from two to four different signals were used: hand raising, flag raising, finished work turned over on a desk, hands folded or a bakery type number system. Hand, and to some extent, flag raising are fairly easy for the observer to see and record, but it is hard to record both together. Other kinds of signals used are almost impossible to observe accurately if they occur in conjunction with another system.

Procedure

The instrument was administered twice to all second grade classrooms in the four established sites. One classroom at each site was

randomly chosen to be observed twice for each administration. The first administration, the field test, was done over a one-month period by one observer. The second administration, the final pass, was done by eight observers (two for each site) over a two-week period.

In accordance with step 20 of Figure 1, a training-program to teach the administrators to use the instrument was developed (Leinhardt, 1972). The program consists of training for knowledge of the instructions, ability to unitize, and ability to categorize teachers' verbal interactions. In-class training focused on recording the distribution of teachers' interactions and record collection.

By the completion of the second pass, data had been obtained on 30 classrooms. On twenty-one classrooms there were two sets of measures and on eight there were four sets. This provided enough data to estimate the reliability of the instrument. The next section examines the problem of observational reliability, followed by discussions of the inter-observer reliability, short-term teacher stability, and long-term reliability of the instrument.

Reliability

The concepts of reliability and validity involve procedures by which confidence in a measuring device may be established. They lend support to the assertion that the measure consistently reports the same situation the same way, and that the measure actually represents that which it is supposed to be measuring. In our case, the major challenges to reliability were that different observers regarded the same event differently, and the lack of stability or representativeness of the behavior observed. The domains of particular interest here are the inter-observer reliability and the stability of teacher behaviors over time.

Several procedures are available for calculating either the overall reliability of data or some specific aspect of reliability, such as inter-observer agreement. No one of them is completely appropriate to the problem of calculating the reliability of our data. In keeping with the spirit, if not the specific method, of Cronbach's "Theory of Generalizability: A Liberalization of Reliability Theory" (Cronbach, Rajaratnam, & Gleser, 1963) where " 'Reliability Theory' is interpreted as a theory regarding the adequacy with which one can generalize from one observation to a universe of observations" (p. 137), I will present a variety of evidence, some of which used traditional estimations of reliability coefficients and others did not, to support the generalizability, or lack of it, for this data set.

Three aspects of reliability were estimated: inter-observer, short-term stability of the teacher behavior, and long-term reliability. The results are reported in Tables 1, 2, 3, and 4. The reasons for seeking more complete information about the reliability of the instrument go back to the initial points mentioned in this paper--there is a very definite need that the data which represent measures of implementation be credible. One way of establishing such credibility is to show that the data are reliable and valid.

Inter-Observer Reliability. Whenever human observation is a basis for measurement, one is faced with the problem of individual differences in observers producing differing results, when in fact, they should have produced the same results. If one is dealing with several observers, the problem is to get all the observers to code the same event in the same way. Table 1 presents the in-class inter-observer agreement for nine observers. Agreement was checked by taking a ratio for each category between two recordings of one situation at one time. All of the observers had an agreement check with me, and in those cases where

TABLE 1

The Ratio of Inter-observer Agreements by Category

	Developer & Observer A	Developer & Observer B	Developer & Observer C	Observers B & C	Developer & Observer D	Developer & Observer E	Observers E & D	Developer & Observer F	Developer & Observer G	Observers F & G	Developer & Observer H	(Mean) Ave. by Category
Total	97	96	100	96	90	90	100	81	90	84	100	93
Cognitive	100	100	82	82	92	100	92	75	75	100	100	98
Management	96	81	7	63	63	71	88	65	65	90	100	78
Percent Negative	100*	75	75	100	100	100	100	100	74	74	100*	90
Distribution Total	95	79	80	64	80	64	80	72	90	80	100	80
Distribution Cognitive	74	86	99	87	98	84	82	46	41	95	100	81
Distribution Management	98	78	33	26	43	78	55	25	31	80	100	59
Average by Individual	94	85	78	74	81	84	85	66	67	86	100	82

*No negatives were observed.

two people were trained together, the ratio of agreement between them is also given. This table presents the range over individuals and categories, and how much each individual agrees on each category. The overall reliability for observers is 82 percent. The range across observers is 66 to 100 percent and across categories is 29 to 98 percent. Because the category distribution of management contacts had such a low reliability, it was dropped.

Short-term Stability. Short-term stability refers to the stability of a teacher's observed behavior over approximately 48 hours. This was checked by having one randomly selected classroom at each site observed twice in two days. The reason for calculating short-term stability is to show that the observed teacher characteristics remain relatively stable over a short period of time.

A summary of short-term stability estimates for several variables is given in Table 2. The variables chosen were ones which would vary over a 48-hour period (or less). The observers were not instructed to recount enrollment, recheck the number of years of experience, or record other context data. Therefore, these variables are not included in the estimations.

The estimates given in Table 2 were obtained from a two-way mixed-model (rows [teachers] random, columns [time] fixed) repeated measures ANOVA of teachers by time for each category for each pass. (That is, separate ANOVA's were calculated for each pass and each category.) From this, two estimates can be calculated which account for the variance due to teachers. One is Cronbach's estimate (which is not presented), (Cronbach, 1971); the other is an eta squared. The eta is obtained by dividing the sums of squares due to teachers by the total sums of squares (i.e., η^2 SSTeachers/SSTotal). This estimate gives the

TABLE 2

Estimates of Short-Term Stability of Teacher Behaviors

Variable	1st Pass	2nd Pass
Percent of children present	.80	.78
Number of days since the last test	.99	.94
Number of cognitive statements	.82	.87
Number of management statements	.80	.96
Distribution of cognitive statements	.83	.065
Percent of unique assignments	.50	.85
Percent of negative statements	.37	.98

amount of variance explained by having the same teacher versus different teachers observed. Table 2 shows the results of estimating short-term reliability by an eta squared. (Both estimates were very close; for further discussion of a comparison of the two estimates, see Leinhardt, 1972.) The average stability for pass one is .73, and .78 for pass two.

The extremely low stability (.065) associated with the distribution of cognitive contacts on the second pass is due largely to one classroom. As shown in Table 3, a classroom at Site 1 had an extremely low distribution score on the first observation and an extremely high score on the second. On the second observation of this classroom, the teacher was planning a field trip for the day but kept the students in school one extra hour so that the observation could be made--a fact unknown to us at the time. Almost no classroom work went on during the observational period; there was a high degree of disruptive behavior. Due to the low stability of the distribution of cognitives in the classroom for the second pass, the measure was substituted in the final data set by the measures from the first set.

Long-Term Reliability. Long-term reliability refers, in this case, to the consistency over one and one-half months of some of the measures which remained unchanged from the field test and final version. The purpose of estimating this is in part the same as estimating short-term stability. But it serves the additional function of estimating the reliability of the instrument in recording some events which presumably do not change. The reliability estimates are reported in Table 4.

The stability of the following measures is reasonably high: teacher experience, enrollment, ratio of boys to girls, and sequence of exploratory--while the stability of the remaining measures is low. The consequence of the low reliability is that the two variables will not be used

TABLE 3

Distribution of Cognitive Contacts: Raw Scores for Doublechecks

Site	1st pass		2nd pass	
	Time 1	Time 2	Time 1	Time 2
1	5.33	4.03	2.43	20.79
2	8.06	6.94	4.04	7.27
3	25.36	15.14	3.4	4.7
4	6.08	8.45	7.7	2.4

Note: The classrooms for each site differ from the first pass to the second.

TABLE 4

Estimation of the Long-Term Reliability

Variable	r_{12}
Years of experience	.92
Enrollment	.70
Ratio of boys to girls	.80
Percent of children present	.34
Square feet per pupil	-.13
Sequence of exploratory	.84

to predict performance, although the attendance variable will be averaged across times and used for descriptive purposes.

In addition to the concern about the long-term stability of the variables, there is a concern about the type of information gained between the first pass and the second, or the difference between sending either one or eight people to collect data. One expectation concerning the difference between the groups would be a greater overall variance for each variable on the second pass than on the first, attributable to individual differences between raters. A test was made for the assumption of equality of overall dispersion. The variance-covariance matrixes from the two passes were found to come from similar populations. That is, the null hypothesis H_o: $D_1 = D_2 = \Delta$ is retained ($F = .921$). This is not, therefore, an estimate of the reliability between one observer and eight observers, but rather, evidence for the assertion that the two situations were comparable in variance and covariance.

Validity

There are two challenges to the validity of this instrument. First, it is possible that the characteristics examined are measured accurately, but are not the relevant ones in terms of the final performance of a class. Second, the characteristics selected may be the most important ones, but the manner of measuring them is not sensitive enough to reveal significant (in the sense of useful) differences. The first challenge has been discussed in the presentation of the instrument. The second challenge was answered in two steps. First, intercorrelations between the input, process, and output variables were examined to determine if the relationships among the variables were consistent with the theoretical basis of the model (they were). Then, the relationships between the output and process variables were examined controlling for the input variables

(e.g., a partial correlation) (Leinhardt, 1972). A selected group of the residual process variables (to be discussed later) accounted for 46 percent of the residual output variance (Cooley & Leinhardt, 1973). The instrument appears to be both sensitive to differences between classrooms and useful in explaining outcomes in terms of achievement scores.

Findings

The purpose of developing the implementation instrument was to provide a source of information which could broaden the interpretation of achievement data. The specific information obtained from this can be examined in terms of three questions. First, which changes in variables in the classroom appear to affect the achievement of students? Second, what differences are observed between the implementation of the program in the field versus its implementation in laboratory classrooms? Third, how is the educational model transformed when it is implemented?

In considering the relationship between classroom variables and achievement variables, it is crucial first to consider how initial input, process measures, and output measures relate to each other. It is obvious that one could obtain strong relationships between what appeared to be measures of process and output by merely having the process measures be surrogate input measures. For example, teacher experience with our program correlates .47 with student achievement, but it also correlates .58 with IQ. When IQ is partialed out of both achievement and teacher experience, the partial correlation between experience and achievement goes to .05. That is, in this case teacher experience is confounded with input variables. In this data set the IQ means (considered input) correlate .68 with arithmetic means which leaves 54 percent of the variance in achievement to be explained by measures of the classroom which are not themselves accidental measures of input.

Table 5 shows the partial correlation between five classroom residuals and Wide Range Achievement Math Residuals. The variables which are negatively related to achievement when IQ is controlled are a large class, more boys than girls, and a larger number of negative contacts. Only the last variable is one which relates specifically to the LRDC program. That is, in the LRDC model, the emphasis is on having the teacher reinforce learning behaviors rather than punish inappropriate behavior. The variables which are positively correlated with achievement are the number of days between tests and the amount of class time spent on mathematics. The second finding is predictable; although it can influence decisions about amounts of time devoted to any one curriculum area, it does not significantly influence the model. The finding that higher achievement is associated with greater time between tests is both startling and intriguing. This would not be expected by the model and it poses some interesting possibilities. Perhaps teachers who frequently test spend less class time tutoring or teaching or adapting assignments to meet individual needs. Or perhaps frequent testing of a child in and of itself is dysfunctional to increases in learning. In either case, it is the type of information which is important to verify and feed back to developers and implementors.

In addition to its analytic function, the instrument was to provide information about the implementation of the program in the field and to permit a comparison of the implementation between the field and the laboratory sites. It would be impossible within the scope of this paper to present a detailed description of each classroom on each variable. Instead, two descriptions will be provided: first, a comparison between the field and laboratory sites and, second, a general discussion of how the model looks in all the classrooms examined.

TABLE 5

Partial Correlations Controlling for Entering Abilities

Class enrollment	1.00	-.27	-.34	.17	.11	-.30
Ratio of boys to girls	-.27	1.00	-.22	.22	-.36	-30
Time between tests	-.34	-.22	-1.00	.00	.00	.32
Percentage of negative contacts	.17	.22	00	1.00	-.19	-.37
Amount of time spent on mathematics	.11	-.36	00	-.19	1.00	.47
WRAT math means	-.30	-.30	.32	-.37	.47	1.00

Table 6 shows the site averages on seventeen variables. In comparing the laboratory and field sites, those variables which appear similar in both field sites and developmental sites will be discussed first, followed by those variables for which there is a difference between the developmental and field sites.

The variables on which there is similarity are teacher experience with the LRDC Model, sex ratio in the class, percentage of unique assignments, percentage of negative contacts, access to play following work, the number of adults traveling, and the number of minutes of math or reading per day. For most of the variables, the field and developmental sites look similar.

The variables on which there is a difference, however, are quite interesting. The number of pupils per class, especially when considered in light of the percentage of children present is smaller in the developmental schools than in the field site schools. This is an important difference when examining outcome measures and per pupil expenditures. There are other differences; in general, more time between tests elapses for those in developmental schools, they make fewer cognitive contacts and more management contacts, and the teacher attention was distributed more evenly. The converse of all these findings is true for the field sites. It would appear that in some respects, the field schools perform in a way that more closely resembles the expressed model than do the developmental schools. A possible explanation of this is that developers are constantly changing the model in the schools to which they have access, placing demands on the teachers so that the classrooms do not strictly follow written or spoken guidelines.

A more relevant question than the comparison of field and developmental schools is how much do the schools look like the model? The

Table 6

Means and Standard Deviations by Sites of 30 Classrooms on 17 Process Variables

Site	Experience		Class Size		Percent Present		Ratio Boys to Girls		Percent Unique		Average Number of Days Since Last Test	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
1	0.2	0.44	27.4	2.41	89.0	5	1.10	.13	57	08	09.06	9.40
2	1.8	0.41	24.0	1.79	96.6	1	1.17	.25	51	04	06.85	2.61
3	1.3	1.03	21.0	1.90	92.5	5	1.18	.45	55	08	07.26	4.50
4	0.0	0.00	25.0	0.63	87.8	7	1.04	.24	56	18	03.15	1.33
Developmental 1	0.4	0.55	22.6	3.20	87.8	9	0.78	.21	48	10	16.72	6.10
Developmental 2	6.0	2.83	22.5	0.70	84.5	14	1.15	.21	60	05	10.68	4.09

Site	Number of Cognitive Contacts		Number of Management Contacts		Percent Negative Contacts		Distribution Total		Distribution (Cognitive)		Play Follows Work	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
1	13.40	05.98	20.20	06.30	6	4	15.83	08.93	07.60	2.70	0.40	0.55
2	19.02	07.50	10.56	03.40	1	2	18.75	04.26	10.92	3.20	1.00	0.00
3	28.00	10.75	15.30	09.80	3	3	20.88	11.43	09.71	3.75	0.00	0.00
4	16.08	04.54	14.25	04.40	6	4	20.33	12.31	08.75	4.94	0.33	0.41
Developmental 1	11.04	03.20	31.00	19.33	2	2	11.87	10.65	08.74	6.60	0.20	0.45
Developmental 2	16.50	10.60	19.50	09.19	5	7	04.45	00.35	05.73	0.03	0.50	0.07

Site	Number of Adults Traveling		Child Obtains Own Work		Number of Minutes of Math per Day		Number of Minutes of Reading per Day		Number of Days the Teacher was Absent	
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
1	1.80	0.45	1.00	0.00	45.00	0.00	90.0	0.00	6.16	4.35
2	2.16	0.41	0.33	0.52	60.00	0.00	120.0	0.00	5.16	3.87
3	1.50	0.83	0.00	0.00	44.16	5.85	88.3	4.08	6.42	3.80
4	2.30	0.82	0.00	0.00	60.00	0.00	120.0	0.00	8.25	2.22
Developmental 1	1.40	0.55	0.00	0.00	60.00	0.00	90.0	0.00	-----	----
Developmental 2	2.00	0.00	0.00	0.00	45.00	0.00	90.0	0.00	7.50	7.70

answer cannot be given a precise value, but some general observations can be made. If one observes the Follow Through classrooms using the LRDC program to determine if they operate in a manner similar to traditional classrooms, the answer is clearly no. All of the classrooms visited are individualized to some extent. Some sites have modified the model in a specific way, such as having two teachers travel, others have extended the management system from the LRDC program curriculum area to all curriculum areas. Some classrooms test frequently, but tend to give children similar assignments rather than individual ones. No classroom assigns just those pages needed by the child as indicated by the test; rather, most of them start at the first page a child needs as indicated by a test and assigns a block of pages after that. There are at least four different styles of traveling, all of which are compatible with the LRDC Instructional Model.

What is clear from Table 6 is that the program can be implemented in diverse settings. It is also clear that the program undergoes a certain amount of modification in the field. One of the questions raised by a study like this is what modifications in the model, made in the field, improve the model. For example, perhaps the model should place less emphasis on frequent testing and unique prescription and more emphasis on differing modes of transmitting information. In a well sequenced curriculum it may not be necessary to continuously monitor progress; and the same effort may be better expended to provide a diversity of curriculum objectives and a means to meet them.

Limitations

The instrument does not provide information on all of the domains initially identified. Some variables were very difficult to measure without extensive clinical data or without developing separate measurement

instruments for them (e.g., student independence). Some variables were lost because of the inability to obtain the information in a reliable fashion.[4] Still other variables were omitted because there appeared to be no measured difference between classrooms. However, a good start has been made in the construction of a reliable and valid instrument for measuring classroom process and the implementation of the LRDC Instructional Model.

Implications

An implementation instrument provides information about the educational processes that occur in classrooms using a particular innovation. The instrument can be a valuable tool to evaluators examining the overall results of an innovation in explaining those results, but it is also a useful tool for implementors and developers. For implementors, it provides information about the success of the implementation relative to the model. For developers, the instrument can provide information on the consequences, both positive and negative, of unintended changes in the model. This information in turn can become the basis for change in specific programs and overall assumptions of an educational model.

References

Campbell, D. Reforms as experiments. American Psychologist, 1969, 24(4), 409-429.

Cooley, W. Methods of evaluating school innovations. Invited address of the 79th Annual Convention of the American Psychological Association, Washington, D. C., September 1971.

Cooley, W., & Leinhardt, G. Evaluating individualized education in the elementary school. Paper presented at the Banff International Conference on Behavior Modification, March 1973.

Cronbach, L. J. Test validation. In R. Thorndike (Ed.), Educational Measurement. (2nd ed.) Washington, D. C.: American Council on Education, 1971.

Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. Theory of generalizability: A liberalization of reliability theory. The British Journal of Statistical Psychology, 1963, 16(2).

Ebel, R. Estimation of the reliability ratings. Psychometrika, 1951, 16(4), 407-424.

Gagné, R. M. The conditions of learning. New York: Holt, Rinehart & Winston, 1970.

Glaser, R. Evaluation of instruction and changing educational models. Los Angeles: Center for the Study of Evaluation of Instructional Programs, 1968. (Also LRDC Reprint 46.)

Leinhardt, G. The boojum of evaluation: Implementation, some measures. Unpublished manuscript, University of Pittsburgh, Learning Research and Development Center, 1972.

Lindvall, C. M., & Cox, R. Evaluation as a tool in curriculum development: The IPI evaluation program. AERA Series on Curriculum Development, No. 5. Chicago: Rand McNally, 1970.

McGraw, B., Wardrop, J., & Bunda, M. J. Classroom observations schemes: Where are the errors? American Educational Research Journal, 1972, 9(1), 13-27.

Medley, D., & Metzel, H. Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.), Handbook of research in teaching. Chicago: Rand McNally, 1961. Pp. 274-328.

Reynolds, L., Light, J., & Mueller, F. The effects of reinforcing quality or quantity of academic performance. Paper presented at the Annual Meeting of the American Educational Research Assocation, New York, 1971.

Rosenshine, B. Evaluation of instruction. Review of Educational Research, 1970, 40, 279-330.

Rohwer, W. Prime time for education: Early childhood or adolescence? Harvard Educational Review, 1971, 41(3), 335-341.

Scriven, M. Curriculum evaluation: Potential and reality. In J. Weiss (Ed.), Curriculum theory network. Toronto: Ontario Institute for Studies in Education, 1972. Pp. 182-192.

Simon, A., & Boyer, G. Mirrors for behavior. Vols. 1-16. Philadelphia: Research for Better Schools, 1967, 1970.

Stake, R. E. Toward a technology for the evaluation of educational programs. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), Perspectives of Curriculum Evaluation. AERA Series on Curriculum Evaluation, No. 1. Chicago: Rand McNally, 1967.

Stake, R. E. The decision: Does classroom observation belong in an evaluation plan? In J. Gallagher, G. Nuthall, & B. Rosenshine (Eds.), AERA Monograph Series on Curriculum Evaluation, No. 6. Chicago: Rand McNally, 1971.

Stufflebeam, D. The use of experimental design in educational evaluation. Journal of Educational Measurement, 1971, 8(4).

THE USE OF OBSERVATIONAL DATA FOR FORMATIVE EVALUATION OF AN INSTRUCTIONAL MODEL[1]

Margaret C. Wang

Learning Research and Development Center
University of Pittsburgh

This paper discusses the critical functions classroom observational techniques can play in the formative evaluation of an instructional model. The study reported in the paper is part of an ongoing series of studies designed to develop techniques and procedures to obtain information for formative evaluation of an early learning environment, designed to adapt to the individual learning characteristics of the student. Specifically, the study was designed to gather descriptive data in classrooms settings, under two different instructional-learning management systems designed to implement the LRDC Instructional Model (see Appendix A) in the early learning grades. Our aim was to identify techniques and measures that are effective in obtaining information about the differential effects of instructional-learning management systems on the instructional-learning processes and student learning outcomes.

The techniques and measures developed for the study included the use of systematic observational schedules designed to gather information

---

[1] The research reported herein was supported in part by a grant from the Ford Foundation, and in part by a grant from the United States Office of Education to the Learning Research and Development Center. The opinions expressed do not necessarily reflect the position or policy of the sponsoring agencies and no official endorsement should be inferred.

concerning classroom instructional-learning processes of the student, as well as an observational technique designed to obtain detailed narrative descriptions of the learning environment and classroom behaviors.

The Setting

The study was carried out in a first grade classroom of an inner-city public elementary school, serving as one of the LRDC developmental schools. The first grade, operating under the LRDC Instructional Model, included two major components, the "prescriptive" learning component and the "extended" learning component. The prescriptive learning component of the program included math, reading, and perceptual skills. In this component the students were assigned to learning activities on the basis of formal diagnostic test results. Extended learning activities encouraged student initiated activities. These generally included more open-ended projects in such subject matter areas as math, science, social studies, reading, writing, prereading, and language arts; in addition, they included a variety of creative arts activities, construction activities, and conceptual games.

Two prototype instructional-learning management systems were designed to implement the LRDC Instructional Model in school settings. The management systems were the Block Schedule System and the Self-Schedule System. Both systems were designed to implement all aspects of the LRDC Instructional Model for early learning grades (Wang, Mazza, Leinhardt, & Millmore, 1971). Under the Block Schedule System, the school day is divided into time blocks. In general, the class begins with block periods for "prescriptive" learning, followed by a block period for "extended" learning. For first grade, the prescriptive learning block periods are reading, math, and perceptual skills. The extended learning

period is usually scheduled at the end of the school day for those who complete all the prescriptive assignments for the day.

Under the Self-Schedule System, as in the Block Schedule System, the student, based on the diagnostic test results, receives daily assignments (prescriptions) in reading, math, and perceptual skills. However, no specific time block is assigned for a specific type of learning activity. The students can work on the learning activities prescribed by the teacher or on any extended learning tasks of their own choice at any time. Under the Self-Schedule System, the students are given the opportunity to make their own decisions on when they will do what, with the exception that some parts of the what are prescribed by the teacher.

Method

Design. A repeated-measures design was used for this study. The Block Schedule System was used during the first four months of the school year (considered as the baseline period [B_1]), although actual data collection for this period did not begin until the third month. The Self-Schedule System was used during the four months immediately following B_1 (considered as the experimental [E]). The Block Schedule System was used again during the reversal period (B_2), which lasted one month immediately following the experimental period.

Sample. All the students of the first grade class selected for this pilot study served as subjects. The total number of students enrolled in the class was 28. The mean chronological age of the students during the experimental period was six years and seven months. The mean IQ for the group was 97, and the standard deviation was 11.7. The students came from an inner-city neighborhood, with the majority of them from low income families.

Data-gathering techniques and measures. A systematic observational schedule was designed to gather information about student learning processes and patterns of interactions in classroom settings operating under the LRDC Instructional Model. The Student Behavior Observation Schedule (SBOS) was designed to obtain answers to the following questions (see Appendix D for an example of the SBOS):

1. What type of activity is the student engaging in (prescriptive or exploratory)?
2. Under what conditions does he choose to work:
   a. Does he work with others interactively?
   b. Does he work on tasks in isolation?
3. Does he complete the activity he chose to do first or does he switch from one activity to another without completing the initial one?
4. How did he decide on what he wants to do--assigned by the teacher, by invitation from another student, or self invitation?
5. What is the major tone or the manner in which the task was carried out--aimless, purposeful, or inattentive?
6. What is the predominant mode and purpose of his interactions with the teacher?
7. What is the predominant mode and purpose of his interactions with other students in his class?

The SBOS was designed to code ongoing student classroom behaviors. The schedule consists of one recording sheet, which includes seven major behavior categories, and spaces for recording five time samples of behaviors of a single student. An observer instructional manual, which includes definitions of the major behavior categories and the behavioral subitems listed under each category, was also developed for training purposes (Wang, 1972).

The technique of narrative recording of student behaviors was also used to gather information for the study. The narrative recordings were used: (1) to study, in depth, the differential effects the different management model has on individual students of certain known unique characteristics; (2) to find out whether the systematic observational schedule alone is adequate to provide the desired information; (3) to compare the observational data obtained from the SBOS with the data obtained from the more comprehensive and more detailed narrative recordings of behavior; and (4) to determine whether the narrative recordings indeed contribute additional pertinent information about student learning that cannot be obtained from the SBOS.

The narrative recording of student behavior focused on student behavior relevant to two major questions:

1. What are some of the noticeable differences in the student's behavior that can be attributed to the management system?
   a. Choice of activity.
   b. How chosen?
   c. Whom he/she chose to work with.
   d. Manner in which activity is carried out.
   e. Predominant mode of interaction.

2. What are some of the learning outcomes produced by this particular management system?
   a. Student's ability to make choices.
   b. Student's ability to plan his day.
   c. Student's ability to meet his curricular demands.
   d. Social interaction with peers.
   e. Particular achievement in prescriptive and extended learning.
   f. Attending behavior.

In addition to the observational data, we also kept a record of the total number of prescriptive learning and extended learning tasks completed during each school day for each observational cycle of the study.

Procedure. Each student in the class was observed for 15 minutes at a time for a total of 75 minutes during B_1, E, and B_2. A specific observational schedule was established to insure that a 15-minute observation out of each different hourly segment of the school day would be made for each student. The school day was divided into one-hour segments. The 75 minutes of observation for each student occurred over two to three days and in no case was a given student observed twice within two consecutive hourly segments. Under this stipulation, four different students were observed during each hour. The students were ranked alphabetically to establish the order of the observation sequence. The observer followed the established observation sequence as closely as possible. Occasionally, because of absences and other unexpected events, some minor modification in the observational sequence was made.

The observer focused on one student at a time. A stopwatch was used to time the observation intervals, and a frequency coding method was used to record behavior occurrence. The 15-minute observation of each student was broken down into a minute of observation, during which the observer coded whatever behaviors she observed occurring under the two interactional categories as listed in the SBOS (categories 1 and 2), followed by a two-minute rest period, during which the observer made appropriate entries for behaviors listed under other categories (3 through 7). Actual observation time during each 15-minute observation period was on minutes 1, 4, 7, 10, and 13, a total of five three-minute observation intervals. At the end of the 15-minute observation of Student 1, the observer began a 15-minute observation of Student 2, following the

same procedures. The observer used a new recording sheet for each new 15-minute observation.

In addition to the observations made for every student included in the sample, a group of six subjects from the sample were selected for intensive study. These six children were selected on the basis of their learning progress in prescriptive curricula and teacher judgment. One girl and one boy were selected from each of the student achievement categories--advanced, average, and below average achievement. Narrative description of student behaviors, as well as the SBOS, were used to obtain information about the six students in the intensive sample.

During each of the three observational cycles (B_1, E, and B_2), in addition to the data obtained from the SBOS for each student included in the intensive sample, narrative description records of a 15-minute observation period per day were made for five consecutive days. Therefore, counting both types of observations, we have collected a total of two and one-half hours of observational data for each student included in the intensive sample. To insure that the time samples for any given student were taken from different times of the school day, a schedule was designed to systematically alter the time of the day during which the additional 15-minute observation for each student was made.

During each 15-minute observation time segment, one child was observed at a time; the observer first used the SBOS to code behaviors as they occurred during the first minute of each three-minute interval. (The procedure used during the first minute of observation is the same as the procedure outlined for obtaining data from the SBOS for the total sample.) During the second and third minutes of each time interval, the observer, using the preplanned guidelines for the narrative description of the student behaviors and learning conditions, wrote down all the

pertinent facts for the first minute, using an improvised shorthand method developed by the observer. In the same fashion, the second observation interval began on minute 4. The same procedure is repeated for the remaining intervals on minutes 7, 10, and 13. After each 15-minute observation of a given student, the observer used as much time as she needed to add to the narrative description any behaviors and facts that occurred during each of the five observation intervals. The same procedure was followed for each 15-minute observation for the entire intensive study sample.

Data Analyses

Observational data obtained from the SBOS were analyzed under two main classifications: the student learning process and the student interaction patterns. Observational results of student learning processes are summarized in Table 1.

The inter-observer agreement was highly consistent for all but one behavior item, "group interactive" under the behavior category entitled "setting." (The reliability coefficient was .60.) The low reliability coefficient in this case probably resulted from a combination of the infrequency of occurrence of the behavior, and the difficulty in determining the nature of behavior in the group setting. It was difficult for the observer to get within the close physical range required to determine accurately whether the subject was actually interacting with other students, or just working in a parallel fashion among other students in the same work area.

For each of the observational cycles, Baseline 1 (B_1), Experimental (E), and Baseline 2 (B_2), the mean ratio of behaviors that occurred for each of the behavior subitems was calculated. The results are summarized in columns 2, 3, and 4 of Table 1. The ratio is obtained from the total number of behaviors observed for a given behavior subitem,

Table 1

Student Learning Processes
(Mean ratio of behavior occurrence)**
N = 25

Variables	Inter-observer Reliability	Observation Cycle: Baseline 1 (Block Schedule) Mean	Baseline 1 S.D.	Experimental (Self-Schedule) Mean	Experimental S.D.	Baseline 2 (Block-Schedule) Mean	Baseline 2 S.D.	Significantly Different Pairs of Means: E.B$_1$	E.B$_2$	B$_1$.B$_2$
1. Activity types										
a. Prescriptive	1.00	.87	.16	.66	.12	.75	.20	•	•	•
b. Extended	1.00	.05	.06	.31	.12	.09	.12	•	•	
c. No activity	1.00	.04	.11	.02	.07	.02	.08			
2. Setting										
a. Group parallel	.96	.01	.03	.15	.16	.12	.30	•		•
b. Group interactive	.60	.04	.09	.09	.24	.05	.20			
c. Individual	.98	.93	.12	.76	.36	.78	.39	•	•	•
d. No activity	1.00	.04	.11	.02	.07	.02	.08			
3. Activity outcome										
a. Complete activity	1.00	.07	.12	.06	.07	.02	.08			
b. Leaves task (incomplete)	.83	.04	.09	.00	.10	.03	.10			
4. Manner in which activities are initiated										
a. Assigned	1.00	.94	.07	.12	.12	.64	.15	•	•	•
b. Invitation by another student	1.00	.00		.00		.02	.13			
c. Self-initiated	.85	.04	.01	.87	.21	.31	.14	•	•	•
d. Not sure	.98	.00		.01	.07	.03	.16			
5. Manner in which activities are carried out										
a. Aimless	1.00	.02	.05	.01	.08	.00	.00			
b. Inattentive	.94	.19	.23	.13	.24	.40	.35	•	•	•
c. Purposeful	.92	.57	.16	.81	.09	.46	.15	•	•	•
d. Waiting	.96	.25	.19	.08	.05	.15	.14	•		•

** Mean ratio of behavior occurrence = $\frac{\text{number of behavior occurrence}}{\text{number of minutes of observation}}$

• t test significant at the .05 level

divided by the total number of minutes of observation time. To test the significance of the difference between the behavior measures obtained during each of the observational cycles, a t-test was performed between the means of each pair of observational cycles. The results are reported in the last three columns of Table 1.

As indicated by the results, the different instructional-learning management systems seem to have differential effects on student classroom activities. The data showed that under the Self-Schedule System (E), students worked significantly more on extended tasks of their own choice, and worked more in group settings; the frequency of prescriptive activity decreased significantly, most of their activities being self-selected and self-initiated rather than assigned; and the students were more purposeful and spent significantly less time waiting for intervention from the teacher. However, no significant differences were found in the observed number of activities completed, or in the number of tasks left unfinished. The "no difference" result of these two behavior categories may be attributed to the "unrepresentativeness" of a sampling technique for observing behaviors that generally occur infrequently.

The data on student behaviors also suggested that students, in general, behaved more similarly during the two baseline periods (using the Block Schedule System) than they did during the experimental period when the Self-Schedule System was implemented. For example, students worked significantly more frequently on prescriptive tasks, they were more inattentive and less purposeful when they worked, and they exhibited significantly more "waiting" behavior during both baseline periods than they did during the experimental period.

Patterns of student classroom interactions under the two different instructional-learning management systems are summarized in Table 2. The results are displayed in the same way as Table 1. About two-thirds

Table 2

Student Interaction Pattern
(Mean ratio of behavior occurrence)**
N = 25

Variables	Inter-observer Reliability	Observation Cycle: Baseline 1 (Block Schedule) Mean	Baseline 1 S.D.	Experimental (Self-Schedule) Mean	Experimental S.D.	Baseline 2 (Block Schedule) Mean	Baseline 2 S.D.	Significantly Different Pairs of Means $E.B_1$	$E.B_2$	$B_1.B_2$
Interaction with teacher:	.88	.25	.19	.15	.28	.18	.21			
1. Type:										
a. Student initiated	.93	.14	.12	09.	.03	08	09	•		•
b. Teacher initiated	.87	.10	.08	06	.02	.11	.08	•	•	
c. Verbal	.33	.15	.18	.11	.17.	.07	.14			
d. Non-verbal	.90	.10	.15	.04	.11	.13	.21			
2. Purpose:	.87									
a. Management	.97	.08	.11	05	11	08	.16			
b. Instructional	.89	.19	.14	.14	.10	.12	.15			
c. Personal	.82	.12	.11	.09	.05	.05	.04			
Interaction with peers:	.88	.60	.37	50	.32	.60	.38			
1. Type:										
a. Student initiated	.94	.33	.15	.24	.08	25	.16	•		•
b. Initiated by another student	.87	.33	.20	.26	08	.34	17		•	
c. Verbal	.94	.58	44	.45	48	.50	.50	•		
d. Non-verbal	.48	.17	.22	.05	.13	17	.25			
2. Purpose:										
a. Sharing (ideas, activities, materials, etc.)	.68	.23	.13	.29	.10	.24	.18			
b. Disagreement	.68	.09	.14	.02	.07	.09	16	•	•	
c. Conversation	.66	.37	.07	.24	.11	.31	.16	•	•	•
d. Seek information or help	.75	.03	.07	.01	04	03	10			
e. Assist others	.67	.06	.08	.01	.07	05	.06			

** Mean ratio of behavior occurrence = $\frac{\text{number of behavior occurrence}}{\text{number of minutes of observation}}$

• t-test significant at the .05 level.

of the inter-observer reliability coefficient for student classroom interactions were within a range close to .90 and about one-third of the behavior items showed low consistency between observers. Behavioral items with low inter-observer reliability coefficients were nonverbal interactions with peers, and the behavior subitems under the general heading of purposes for peer interactions. We have reason to believe that the low reliability coefficient of these behavior subitems is largely attributed to the difficulties in getting both observers simultaneously within a close range of the subject being observed. This is a major technical problem one generally encounters when conducting observational studies in natural settings, and this problem is particularly magnified in classroom settings where students are permitted to freely move about the classroom and interact with their peers. According to the feedback information from the observers, the magnitude of this problem decreased when classroom observations for the study were made; it was much easier for one person to get to a vantage position to focus on the subject being observed than to get two persons to approximately the same position in order to observe the same student at the same time for the reliability study.

The data on student interaction patterns were consistent with the other findings on student learning processes. The students behaved differently under the two different management systems. By examining the mean frequency of behavior occurrence under the heading of student-teacher interaction one notices two things: (1) that students interacted with the teacher more often under the Block Schedule System than during the Self-Schedule System, and (2) that teachers initiated significantly more of the interactions during the Block Schedule System. However, no significant differences were found in the purpose of the interactions. The observational data on student interactions with peers showed that students interacted with one another more often under the Block Schedule

system than in the Self-Schedule System. However, it is interesting to note that a large proportion of the peer interactions under the Block Schedule System were "disagreement" and "conversations."

In general, the results suggest that the SBOS seems to be an effective instrument to use in documenting the behavior differences in student learning processes and interaction patterns in classroom settings. Furthermore, the behavior categories included in the SBOS were relevant to the rationale and hypotheses made on the teaching-learning processes under the two different instructional-learning management systems. The data yielded information that not only made analysis of the selected dimensions of student classroom behaviors unique to our program possible but, more importantly, the data provided us with information to carry out formative evaluation needed to improve and redesign the instructional-learning management systems. As an example of the use of empirical data (data obtained from a systematic study such as the present study) for formative evaluation purposes, I shall discuss the result of our analyses on the effectiveness of the two different management systems in solving the problem of student "waiting time."

Among the many technical problems we encountered in implementing the LRDC Instructional Program was the demand on teacher time. The teachers were not able to get to the students fast enough to both answer questions and to check their work at the same time. Under the individualized instructional program, students generally spent a large percentage of class time "waiting" for teachers instead of working on the tasks. The observational data from the present study, as well as results from previous studies, clearly suggest this unique "waiting" phenomenon (Wang, Mazza, Haines, & Johnson, 1972; and Yeager & Lindvall, 1968). This "waiting time" was generally a frustrating experience for teachers and

students. For students, the "waiting time" was tiresome and boring. Students were not only bored from doing "nothing" while they were waiting, but also from doing the same type of tasks for a long period of time (e.g., some students may be required to work on prescriptive tasks from one subject to another during most of the school day). For the teacher, the student "waiting time" was "troublesome" and "hard to manage." The teachers were required to "travel" among the students at a fast rate in order to get to every student who needed help, and also to keep those "waiting" students from disrupting others at work.

The results of this study suggest that the Self-Schedule System is effective in reducing student "waiting time" to a minimum level. As shown in Table 1 (p. 59), the mean ratio of "waiting" behaviors decreased from .25 for B_1 to .08 for E (the t-test was significant at the .05 level). Although the differences of the percent of waiting time between E and B_2 were not statistically significant, the differences were still substantial. Waiting time for B_2 was .15, almost double the amount of waiting time during E. From a formative evaluation point of view, this preliminary set of data suggests that the Self-Schedule System may be more effective than the Block Schedule System in solving the problem of student "waiting time."

Our data for formative evaluation of the instructional management systems were also used to study the relationship between learning processes and learning outcomes. Canonical correlations were performed to investigate the relationship between student classroom behaviors and student learning outcomes under the two different management systems. In this series of analyses, we were particularly interested in examining the relationship between a linear combination of tasks completed (tasks included in both prescriptive and extended learning), and a linear combination of selected sets of classroom behavior measures. The results of

the canonical correlations are shown in Table 3. For each analysis, the variables included in the criterion set are listed under the first column; the variables included in the predictor set are listed under column 2. The canonical R, chi square value, degrees of freedom and the statistical test of significance are reported in columns 3, 4, 5, and 6, respectively.

The criterion set included in Analyses A and B was mean number of tasks completed in prescriptive and extended learning during Baseline 1 and the Experimental periods. The predictors included in the analyses were seven selected behavior measures obtained by the SBOS. These measures included variables that we have suggested as critical learning process variables that affect student learning outcomes under the LRDC Instructional Program: (a) prescriptive activity, (b) extended learning activity, (c) self-initiated activity, (d) purposeful on-task behavior, (e) waiting for teacher intervention, (f) sharing behavior, and (g) conversation with other children.

The rationale for performing this series of canonical correlations is based on our assumption that what the student did and the manner in which he carried out the learning tasks must have direct implications for student learning outcomes. As shown in Table 3, a significant relationship between the task completion rate and observation measures of student learning processes during the Self-Schedule System was found (during the experimental period), while the relationship between what the students were able to do and the process under which the tasks were carried out during the Block Schedule System (during the baseline period) was not significant.

The results suggest that behaviors exhibited by the students during the experimental period seem to relate to students' task completion rate, while students' behaviors under the Block Schedule System were not

Table 3

Canonical Correlations Between Tasks Completed and Selected Number of Behavior Measures
N = 25

Variables		Canonical*			
Criterion Set	Predictor Set	R	Chi Square	d.f.	t-test
Analysis A:					
1. Prescriptive tasks completed during baseline period	Selected classroom behavior measures for the baseline period:	.48	7.29	14	> .05
2. Extended tasks completed during baseline period	1. Prescriptive activity 2. Extended activity 3. Self-initiated activity 4. Purposeful behavior 5. Waiting for teacher intervention 6. Sharing behavior 7. Conversation with other students				
Analysis B:					
1. Prescriptive tasks completed during experimental period	Selected classroom behavior measures for the experimental period:	.72	25.29	14	< .05
2. Extended tasks completed during experimental period	1. Prescriptive activity 2. Extended activity 3. Self-initiated activity 4. Purposeful behavior 5. Waiting for teacher intervention 6. Sharing behavior 7. Conversation with other students				

* Only the first canonical R's are reported in this table.

related to what they did. This interpretation is further supported by the fact that students spent significantly more time under the Block Schedule System doing prescriptive work, without completing a concomitant amount of prescriptive tasks, and the fact that students under the Block Schedule System also spent more time "waiting" for teachers. In other words, the canonical results for Analysis A, listed in Table 3, seem to suggest that student behaviors under the Block Schedule System were not related to their task completion rate. The fact that the process measures obtained for the SBOS have been good predictors of student learning outcomes under the Self-Schedule System, but not related to learning outcomes under the Block Schedule System, further supported the assumptions we have made in the design of the Self-Schedule System. However, because this is a pilot study with a small sample, these interpretations should be treated as tentative.

Another example of formative evaluation questions we asked from the data in the present study was, "Is the SBOS an effective instrument to use for evaluating the learning processes of individual students?" To answer this question, we examined the observational data obtained for each of the six students included in the intensive sample individually. In addition to the observational data, to investigate the learning processes of each of the six students, we also collected information on a selected number of student learning characteristics. The student characteristic measures are summarized in Table 4. The behavior changes of each student observed during each of the observation cycles, using the SBOS, are summarized in Table 5. Behavior items included in Table 5 are behaviors found to be significantly affected by the different management models (significantly different pairs of means as reported in Tables 1 and 2 [pps. 59 and 61, respectively]). On the whole, the data suggested that the different management models had differential effects on student

Table 4

Student Learning Characteristics of the Intensive Study Sample

Student	CA (mos.)	IQ	Sex	Characteristics: Task Completion Rate* Baseline 1 Prescriptive	Baseline 1 Extended	Experimental Prescriptive	Experimental Extended	Baseline 2 Prescriptive	Baseline 2 Extended	Mastery Score Math (Unit)	Mastery Score Reading (Book)	Achievement Score WRAT (Row) Reading	Spelling	Arithmetic
A	79	97	F	8.49	.42	25.28	1.24	12.50	.38	9	3	30	22	23
B	77	98	F	18.66	4.33	42.45	2.91	23.13	2.25	12	9	48	20	25
C	79	89	M	8.00	5.66	19.11	2.62	12.50	1.75	5	2	31	14	20
D	79	89	F	15.8	1.66	39.2	1.51	13.60	1.50	7	10	48	27	23
E	91	102	M	13.0	5.50	24.77	6.90	28.25	4.75	10	10	53	33	27
F	82	88	M	16.17	4.66	28.09	11.34	17.00	3.50	8	3	29	19	21

* Number of tasks complete week.

Table 5

Summary of Student Behavior and Task Completion Rate of the Six Intensive Study Samples

Variables	Student A			Student B			Student C		
	B_1	E	B_2	B_1	E	B_2	B_1	E	B_2
A. Activity type:									
1. Prescriptive	.67	.79	.83	1.00	.65	90	.80	.58	.75
2. Extended	.00	.18	.10	.00	.31	.20	.05	.27	.13
3. Assigned	1.00	.01	.76	1.00	.13	.79	1.00	.13	.75
4. Self-initiated	.00	.97	.24	.00	.87	.10	.00	.80	.25
B. Manner:									
1. Inattentive									
2. Purposeful	.80	.62	.48	.80	.80	.59	.52	.45	.54
3. Waiting	.27	.07	.10	.00	.33	.11	.25	.08	.15
C. Interaction with teacher:									
1. Student initiated	.13	.06	.00	.20	.40	.17	.14	.07	08
2. Teacher initiated	.13	.07	.03	.20	.06	.07	10	.29	.17
D. Interaction with peers:									
1. Student initiated	.27	.52	.24	.50	.12	.14	.38	.47	.67
2. Initiated by other student	.80	.35	.41	.30	.24	.38	.39	.42	.42
3. Disagreement	.13	.03	.07	.30	.01	.07	.30	.25	08
4. Conversation	.60	.45	.31	.10	.21	.34	.59	.49	.75
E. Tasks completed:									
1. Prescriptive	8.49	25.25	12.50	18.66	42.45	23.13	8.00	19.11	12.50
2. Extended	.42	1.25	38	4.33	2.91	2.25	5.66	2.62	1.75

Table 5 con't

Summary of Student Behavior and Task Completion Rate of the Six Intensive Study Samples

	Variables	Student D			Student E			Student F		
		B_1	E	B_2	B_1	E	B_2	B_1	E	B_2
A.	Activity type:									
	1. Prescriptive	.50	.75	.71	1.00	.44	.79	1.00	.76	.69
	2. Extended	.25	.17	.21	.00	.55	.10	.00	.10	.31
	3. Assigned	.75	.10	.75	1.00	.00	.79	1.00	.12	.69
	4. Self-initiated	.25	.80	.13	.00	.95	.21	.00	.86	.31
B.	Manner:									
	1. Inattentive	.85	.78	.67	1.00	.98	.76	1.00	.58	.59
	2. Purposeful	.05	.08	.04	.10	.00	.28	.00	.14	.21
	3. Waiting									
C.	Interaction with teacher:									
	1. Student initiated	.05	.06	.04	.30	.06	.10	1.00	.14	.03
	2. Teacher initiated	.05	.22	.04	.10	.05	.03	.00	.13	.13
D.	Interaction with peers:									
	1. Student initiated	.45	.22	.21	.00	.14	.07	.50	.42	
	2. Initiated by other student	.95	.39	.50	.30	.11	.07	.10	.51	.45
	3. Disagreement	.15	.04	.25	.10	.03	.00	.10	.41	.36
	4. Conversation	.80	.16	.29	.20	.05	.00	.00	.10	
E.	Tasks completed:									
	1. Prescriptive	15.80	39.20	13.60	13.00	24.77	28.25	16.17	28.07	17.00
	2. Extended	1.00	1.51	1.50	5.50	6.90	4.75	4.66	4.34	3.50

behaviors. We were able to obtain a general picture of the difference in behaviors and outcomes for each student. However, differential effects of the different systems on individual students were not as distinct as we had hypothesized. The general group trend of behavior changes under the different systems prevailed.

The Self-Schedule System was consistently effective in promoting the student's rate of prescriptive task completion (see Figure 1), but it did not affect the rate of extended learning task completion (see Figure 2). As shown in Figure 1, every student consistently had a lower rate of prescriptive learning and task completion during B_1 and B_2, and a significantly higher rate during E. Some differential effects on the rate of extended learning task completion of the individual students can be detected in Figure 2. However, because of the small number of extended learning tasks completed by each student during the different observation cycles, the magnitude of the differences was quite small--the differences were not statistically significant. Nevertheless, when one examines Figure 2 closely, one can see the differences in the trend of behavior changes. For example, the effects of the system on Students A and E are different from that of Students B, C, D, and F. Comparing Figure 2 with Figure 1, the different management systems did not seem to have a consistent effect on students' extended learning task completion rate as they had on students' prescriptive learning task completion rate.

As we analyzed the data obtained from the SBOS for the six students included in the intensive study sample, the lack of "depth" in the data became evident. This supported our notion that observational data obtained from a measure such as the SBOS do not yield sufficient information about the individual student in order to carry out formative evaluation of individual learning processes and outcomes.

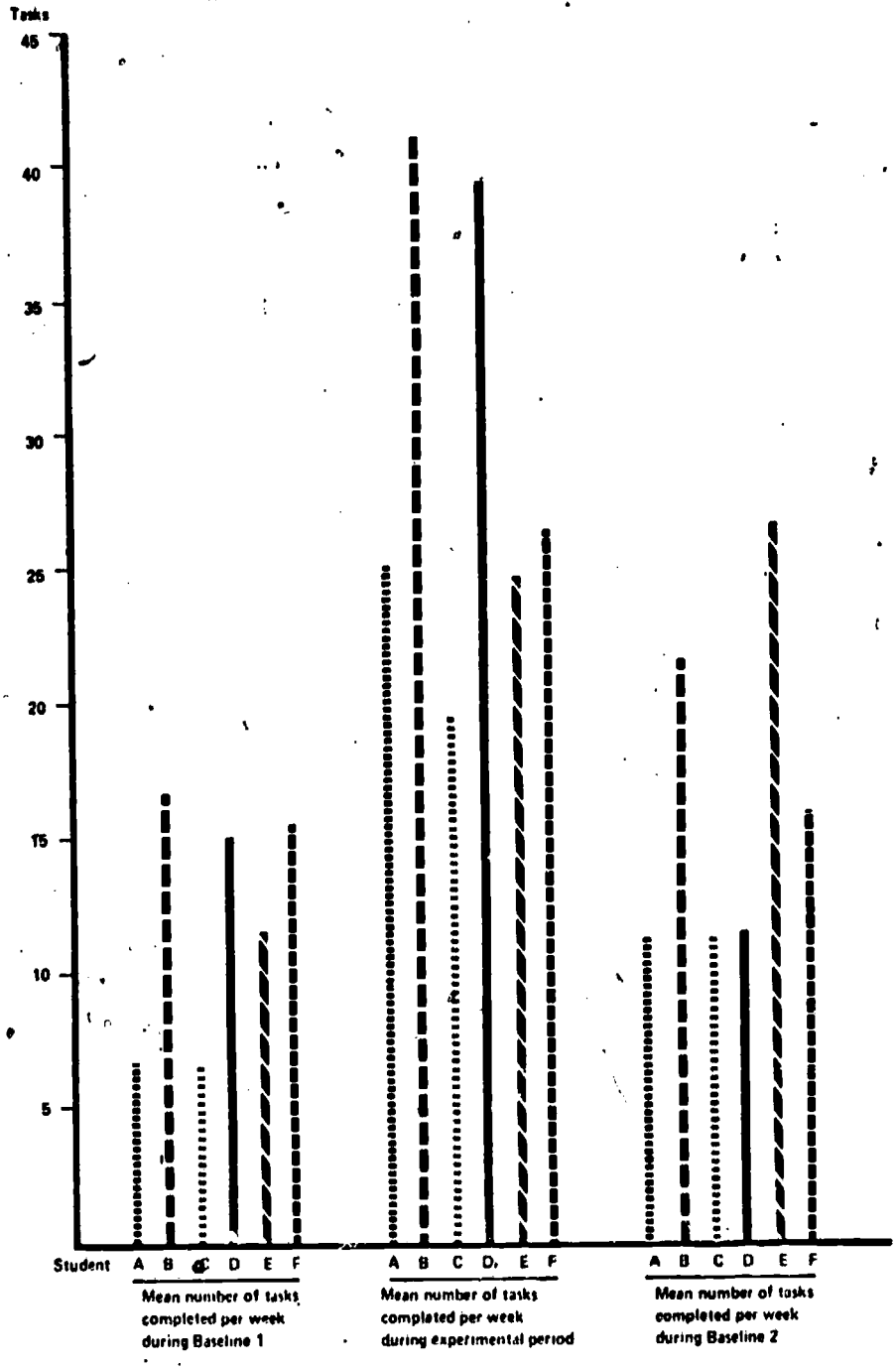


Figure 1: Prescriptive Learning Tasks Completed.

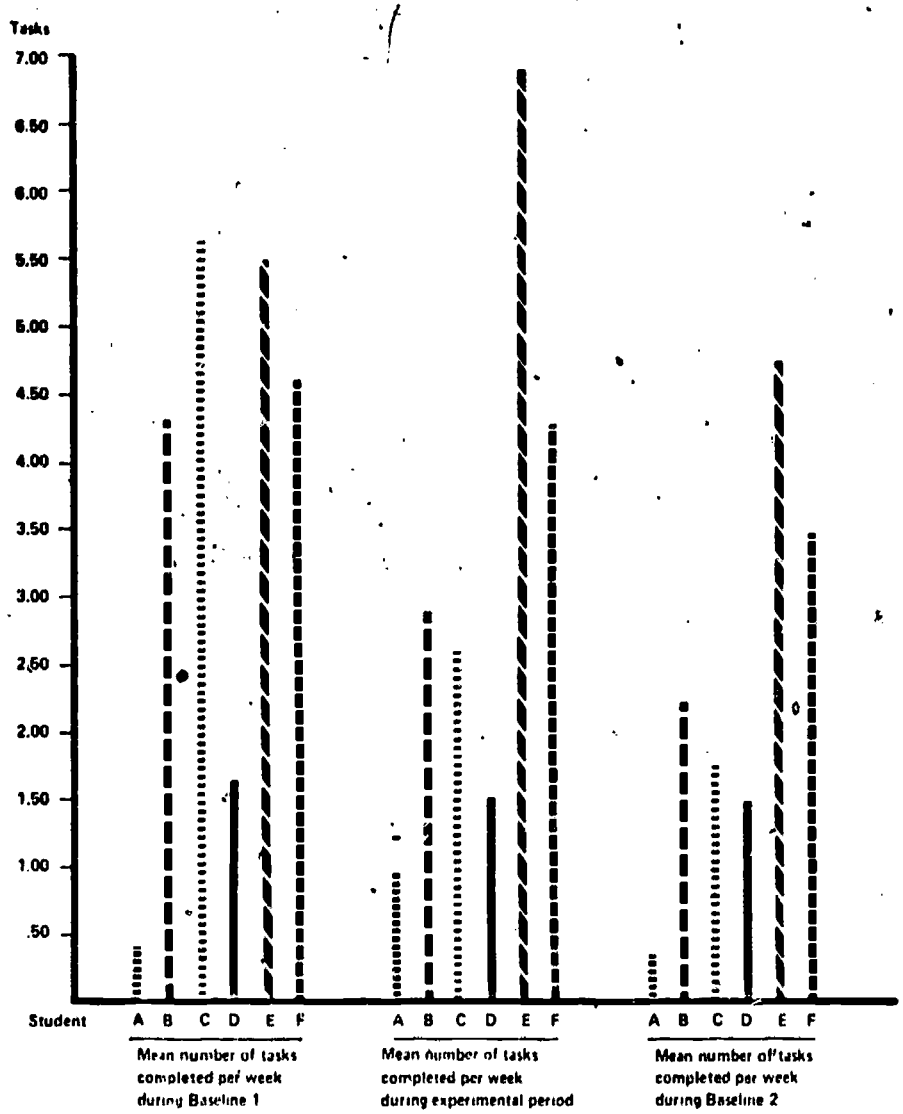


Figure 2: Extended Learning Tasks Completed.

73

To answer the question of whether narrative description data provided additional information for studying, in depth, the differential effects of the management model on the individual student, the combination of observational data obtained from the SBOS and the narrative description data of the six students included in the intensive study sample were analyzed. The narrative records of each student were first compiled and a summary of each student's description was made. The behavior description was then compared with data obtained from the SBOS for each student.

The information obtained from the two different sets of data were found to be in general agreement; there were no specific inconsistencies. However, the data clearly reflected that the information yielded by the two observational techniques served two distinct purposes. The SBOS is effective in formative evaluation of the instructional model, but not sufficient for formative evaluation of the learning processes for individual students. The use of the narrative description method provided more in-depth information about the individual differences in learning processes among students.

Discussion

Through the use of the Student Behavior Observation Schedule (SBOS), we were able to obtain pertinent information about the relationship between the LRDC Instructional Program and its specific effects on student learning processes and student learning outcomes in classroom settings. We found that observational data of this kind is useful in delineating the relationship between what the instructional model requires the teacher to do in the implementation of the LRDC Instructional Program and the effects on the students. We found, for example, that the Self-Schedule System

produced: (1) an increase in the rate of prescriptive task completion per week, although students worked on prescriptive tasks less frequently; (2) more purposeful and attentive behavior on the part of students; (3) a decrease in the frequency of student waiting behavior; (4) an increase in the frequency of students working in group settings; and (5) a decrease in student "disagreement" behavior.

More importantly, the data from the SBOS were useful in giving us the feedback information about the aspects of the instructional-learning management program that did not function as hypothesized. Information of this type is critical to the validation and revision of the design of a management system. The data indicated, for example, that the Self-Schedule System did not effectively increase the rate of task completion in extended learning, even when it was observed that students engaged in extended learning activities more frequently. Further investigation of the results suggests that using the rate of task completion to evaluate extended learning outcomes is misleading. The fact that students did not complete more extended learning tasks (as measured by number of tasks completed) does not necessarily mean that students were less efficient in task completion under the Self-Schedule System. The results may mean that students spent more time working on a given extended learning project, thus completing fewer total tasks. Furthermore, because of the varied nature in the quality of student involvement in the extended learning activities students have chosen to do, comparing outcomes in terms of numbers is neither sufficient nor appropriate. Perhaps using both "length of time per activity" as well as "number of tasks completed" can provide us with a more adequate measure. It is probably more meaningful for a student to spend a relatively longer time to accomplish an extended exploratory project of his choice, than completing more short-term learning projects. Nevertheless, these assumptions will need to

be studied before we can make a definite conclusive statement. The results, however, demonstrated the need to use a combination of many different types of information to conduct formative evaluation of student learning.

We also found that the Self-Schedule System did not increase the frequency of peer interaction as we had hypothesized. This finding implies that a more thorough examination of the type of tasks we have available in the learning environment is necessary. The fact that these students worked significantly more frequently in group parallel settings under the Self-Schedule System suggests that the management system probably was effective in providing the opportunity for the students to interact with one another; but the nature of the tasks available to them may not reinforce and/or permit interaction among the students. This result also illustrates the multiple functions that observational data of this type can serve. The observational data provided us with the formative evaluation information of our instructional-learning content, as well as information about the instructional-learning management system.

Another important input from the observational data concerned the teacher-student interactions. While the frequency of interaction between the teacher and students did not increase, student waiting time decreased. The results can be interpreted as an indicator of the effectiveness of the Self-Schedule System to achieve our objectives--to cut student waiting time to a minumum and to enable the teacher to have a maximum amount of interaction. Although the data do not show an increase in the frequency of teacher-student interaction as we have hypothesized, we have reason to believe that the amount of time per interaction involved has increased under the Self-Schedule System. Another implication that our data suggest is that an increase in the rate of task completion is the result of an increase in the amount of time per interaction, rather than in the frequency

of interactions. A system designed to promote frequency (the rate) of teacher-student interaction, theoretically, should look very different from a system designed to promote extended interaction between the teacher and student.

Summary

This pilot study has shown that classroom observational data can contribute to the formative evaluation of the instructional model, as well as to the formative evaluation of the learning processes and outcomes of individual students. The data obtained for the six students included in the intensive samples provided us with pertinent information about differential effects that the management system can have on individual students. To achieve our goal of providing a truly adaptive learning environment that takes into consideration all aspects of individual student learning needs, information about the effects of certain types of instructional-learning management systems is a prerequisite.

The following summary of interview comments by the teachers who were aware of the type of information we obtained from the observational study illustrates this point:

> The teachers spent quite a bit of time discussing the different ways that kids respond to their learning tasks. They found that some students consistently chose to work on extended learning tasks first, and prescriptive learning tasks later. Others do prescriptive learning tasks first and extended learning tasks later. Still others intersperse the two. They have noticed that students who have the most energy and are most likely to be "messing around" are the ones who spend the most time in extended learning activities. This "energy" dimension seems to cut across academic ability or at least academic standing. They discussed specific students and how they fit into these general patterns. They stated ". . . students who are really task

oriented and are not too fast, like D and E, extended learning used to end at approximately 2:10 or 2:20 and they would barely get into extended learning activities before the day was over. . . . But they were task (prescriptive task) oriented, and they would just do that first . . . R has been doing her two math boxes very quickly in the morning and I will say something like 'Oh, good; now you can get your reading done,' and she will say, 'No, I'm going to extended learning activities now, and then I'm doing my reading because that's the way I'm doing things now.' . . . The students who would spend a lot of time in extended learning cover a range; B and T would happily leap around all day and so would S and D who are at the other end as far as ability and what kind of work they have done. . . . Like T . . . we can see he will get his work done. He will work on two boxes at a time. He is not going to wait for us to come around. . . . That is wasting too much time; he cannot get done fast enough. . . ." The teachers also talked about the importance of interactions among students as an important measure of the success of an activity, in addition to the appropriateness of an activity for a particular student. They expressed their pleasure at seeing M and T working together on building projects. They felt that M was having a unique learning experience building with the much more verbal T. They valued P's interest in listening to the record player, R teaching others how to knit, students playing a card game together, complex block constructions from T, clay creations from M, and sensory play with seeds for T.

It is exciting to know that teachers can have information regarding the student's distinct stylistic differences in the ways they carry out learning tasks. Such information is a useful contribution to work on adapting instructional-learning processes to individual differences. The information forces the teacher to pay attention to the processes used by the individual student, rather than looking only at outcomes.

The results of the study clearly suggested the potential usefulness of the observational techniques in obtaining pertinent evaluation information. The observational data provided us with the information we needed

to investigate the relationship between the instructional-learning system(s) used to implement our instructional program in classroom settings, and what happens to the student learning processes and student learning outcomes. However, as I have pointed out in the discussion of results, a great deal of validation work still needs to be done before wide application of the techniques and measures can be used in classrooms using the LRDC Instructional Program and in other classrooms using different instructional approaches.

As we develop reliable and valid techniques and measures to examine the classroom learning processes and outcomes, and as we are able to obtain more formative evaluation information as I have described, we will be better equipped to design a learning program and learning environment that will be adaptive to the multiplicity of factors and student characteristics that affect student learning in school. It should be our objective, as program developers, to provide teachers with techniques to help them to diagnose individual differences and needs of their students in dimensions other than academic achievement. Developers should also equip teachers with ways to make some of the program objectives more concrete. We will be able to make the difficult job of teaching in an individualized instructional program less difficult and more explicit to the teacher only when we can clearly define the instructional-learning management system, the learning conditions, and what we can expect to observe in the nature of the interrelationship between teacher behavior, the instructional-learning management system, and student behavior.

References

Bolvin, J. O., & Glaser, R. Individualizing instruction. In D. W. Allen & E. Seifman (Eds.), The teacher's handbook. Chicago: Scott Foresman, 1971. Pp. 270-279.

Cooley, W. W., & Lohnes, P. R. Multivariate procedures for the behavioral sciences. New York: Wiley, 1962.

Glaser, R. Adaptive education. Paper presented at the Conference on University Teaching and Learning, McGill University, Montreal, Canada, 1971.

Glaser, R. Adapting the elemeatary school curriculum to individual performance. In Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton: Educational Testing Service, 1968.

Jastak, J. F., & Jastak, S. R. The Wide Range Achievement Test. Wilmington, Delaware: Guidance Associates, 1965.

Landvall, C. M., & Bolvin, J. O. Individually Prescribed Instruction: The Oakleaf project. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1966. Working Paper 8.

Resnick, L. B. Design of an early learning curriculum. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1967. Working Paper 16.

Wang, M. C. An instruction manual for the student observation schedule. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1972. Mimeo.

Wang, M. C. Rationale and design of an early learning environment. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1973. Mimeo.

Wang, M. C., Mazza, M., Haines, J., & Johnson, M. Some measured effects of a classroom management model designed for an individualized early learning curriculum. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, April 1972.

Wang, M. C., Mazza, M., Leinhardt, G., & Millmore, M. Exploratory learning program: A teacher's manual. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1971.

Yeager, J. L., & Lindvall, C. M. Evaluating an instructional innovation through the observation of pupil activities. The High School Journal, 1968, 51(6), 248-253.

LEARNING ACTIVITIES IN INDIVIDUALLY PRESCRIBED INSTRUCTION[1]

Joseph Shimron

Learning Research and Development Center
University of Pittsburgh

A major goal of the Primary Education Project (PEP) and Individually Prescribed Instruction (IPI) programs (see Appendix A) is to adapt the educational system to individual differences. The extent to which this adaptation occurs is under continuous study and evaluation. The major purposes of this pilot study were: (1) to determine whether or not distinctive patterns of behavior exist for students of different learning characteristics (e.g., learning rate), and (2) whether or not and the degree to which the IPI system succeeds in adapting its instructional procedures to these behavior patterns.

Method

Measures of Adaptability. The question posed is how well the IPI system actually adapts to individual differences. In order to measure

---

[1] The research reported herein was supported in part by a grant from the Ford Foundation, and in part by a grant from the United States' Office of Education to the Learning Research and Development Center. The opinions expressed do not necessarily reflect the position or policy of the sponsoring agencies and no official endorsement should be inferred. This work was guided by Dr. Robert Glaser and Dr. Lauren Resnick. Their help is deeply appreciated. The author wishes also to thank the teachers at Frick School, Mary Jane Rooney and Geraldine Henderson, for their permission to observe the work in their classroom and for the help they provided during the study.

adaptability, one can observe rates of progress and mastery time. One can also measure general achievement and changes of aptitudes. These measures are very useful in comparing the academic achievement produced by IPI to that produced by other systems. Both Resnick (1967) and Lindvall and Bolvin (1966) have predicted that under an individualized instructional system, students will show a higher level of achievement compared with students studying under traditional instruction. This gain is attributed to the better match that this system achieves between the curriculum and the students' knowledge, needs, and interests. Nevertheless, measures of academic achievement alone provide no information about the degree to which instructional procedures adjust to individual differences which are expressed in actual classroom behavior.

In this study, it was first assumed that in an adaptive system of education, individual differences along cognitive and noncognitive dimensions could freely develop. The IPI system is not a full range "open" system since children follow a prescribed order of learning units. Nevertheless, the fact that children can complete the units at their own pace, and the fact that they do not behave in a permanent teacher-controlled situation (as in the traditional classroom), make IPI open enough to allow for individual differences to be expressed.

Second, it was assumed that if individual differences are actually expressed, it should be possible to detect them by, for example, observational methods. These observational methods may show how different kinds of students behave and how they interact with their peers, teachers, and other factors in the educational setting.

The question remains as to what aspects of behavior best indicate adaptability. Two assumptions about adaptability were made in this study. The first is that adaptability of the curriculum to individual dif-

ferences can be detected by measuring the students' on-task/off-task behavior. If the curriculum is well matched to individual differences (by the curriculum builder and by the classroom teacher), slow and fast students should not differ in the amount of time they spend in on-task behavior, be that prescribed, mutually agreed upon with the teacher, or self-initiated task.

The second assumption is that adaptability of the teacher to individual needs can be detected by measuring the number of teacher-student interactions. The circumstances under which these interactions take place (e.g., on-task or off-task) is also important. In general, it is assumed that there should not be a difference in the availability of the teacher to either slow or fast students. Occasionally, however, when there is an apparent need to compensate for a disability evidenced in the learning progress of a slow student, it is expected that the teacher will initiate encounters with the student in order to assist his learning efforts and to provide extra emotional support and reinforcement.

From the assumptions made above, it can be seen that characteristics of adaptability can be defined in terms of two dimensions: the task-related dimension and the interpersonal dimension.

Three possible kinds of task-related learning were defined: direct on-task behavior, task-oriented behavior, and off-task behavior. Direct on-task behaviors included activities in which the student seemed to be occupied by his tasks (e.g., reading, writing, game playing, counting). The task-oriented category included activities closely related to on-task behavior though not considered as direct on-task behavior (e.g., waiting, arranging an assignment). Such activities could indicate learning tendency or motivation. Activities in which no relation to task performance was apparent were included in the off-task category.

The second behavioral dimension specified the types of behavior engaged in by the student while performing direct on-task, task-oriented, and off-task activities. Three types were defined: no interpersonal interactions, dyadic interactions (student-student, student-teacher), and group interactions.

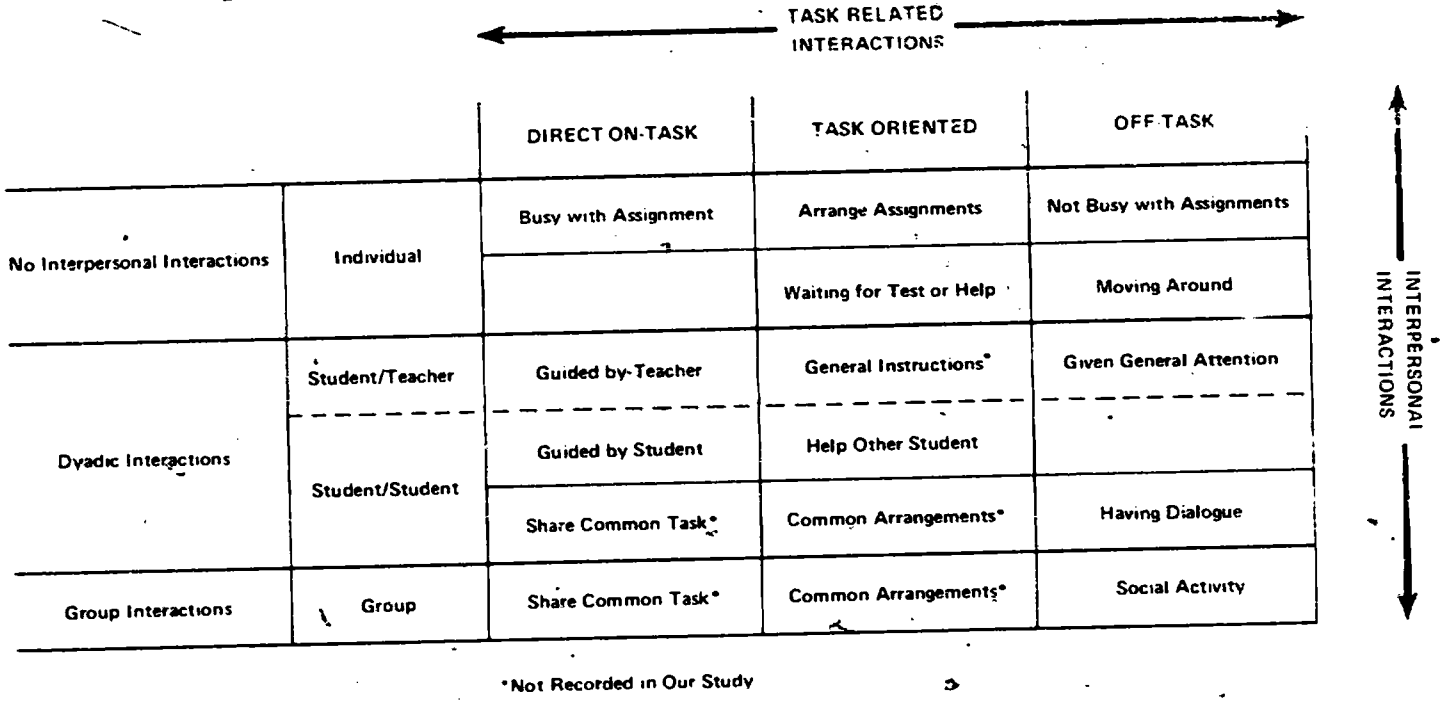
These two dimensions were combined into a matrix from which categories of observed behavior were derived (see Figure 1). Not all cells of the matrix were measured, however, since not all combinations of behavior can occur under the IPI system.

Subjects. The subjects for this pilot study were eight second-grade students from the Frick School, an inner-city school that is associated with the Learning Research and Development Center. Four of the students had mastered the largest number of curricular units and were thus classified as "fast." The other four had mastered the least number of units and were designated as "slow."

Apparatus. In order to record the different amount of time spent in each category of behavior, a graphical recording device with a remote control button box was employed. This device enabled the observer to position himself at any point in the classroom.

Procedure

The observations in this pilot study took place in the winter of 1972 in one of the second-grade classrooms in Frick School. There were 20 students in this classroom. Observations were taken during the morning hours (from 9:30 to 11:30). During these periods, one teacher or aide "traveled" among the students tutoring those who requested help, while the other provided assistance to a student or groups of students for more extended periods of time. Half of the



*Not Recorded in Our Study

Figure 1. Theoretical Structure of an On-Task Interpersonal Categorical System

students in the class worked on math projects and the other half on reading. The next day those who worked on math worked on reading and vice versa. Curriculum units were shelved against the walls and so the students could independently walk to the shelves and pick up his/her appropriate task. The students were also aware of how to arrange the materials for learning, how to reorganize them after they were through and how to restock the units so they were available to others.

Only one student was observed at a time, though each student was observed five times during the study. The percentages of agreement between the two observers were checked by videotape equipment and found to be between 85 and 91 percent. In order to control for subject matter effect, observations were made an equal number of times when the students were studying math and reading. Similarly, the effect of a particular teacher was also controlled. (Two persons, as it is explained above, were guiding students--the main teacher and the aide, although only one was in contact with the student at a time.)

Results and Discussion

The results presented below concentrate on: (1) characteristics of slow and fast students when they worked individually, (2) characteristics of student-teacher interaction of slow and fast students, and (3) characteristics of the peer interactions of slow and fast students.

Individual Domain. Tables 1 and 2 show the results in those categories in which the individual does not interact with either peers or a teacher. On the average, the fast student spent twice as much time working on his assignment than the slow one. The number of occurrences of on-task behavior, however, was about the same. The percentage of time spent in direct off task behavior by slow students was twice that spent by the fast ones. Both the slow and the fast students spent about

Table 1

Mean Percentage of Time Spent in Individual Activities

Activity	Slow Students*		Fast Students*		t-test
	Mean	S.D.	Mean	S.D.	(p. level)
On-task	21.8	3.1	41.5	6.8	.005
Arrange assignment	14.7	6.5	14.2	2.1	N.S.
Waiting for test or help	9.8	7.3	11.7	4.0	N.S.
Off-task (sitting)	22.9	10.2	10.4	8.1	.10
Off-task (moving in class)	3.7	3.0	1.7	0.5	.10

* Four slow students and four fast students were observed. Each student was observed five times during the year for a period of 20 minutes. Total observation time on each student was 100 minutes

Table 2

Mean Number of Occurrences of Individual Activities

Activity	Slow Students*		Fast Students*		t-test
	Mean	S.D.	Mean	S.D.	(p. level)
On-task	8.7	2.2	8.9	2.0	N.S.
Arrange assignment	4.9	1.3	4.5	0.9	N.S.
Waiting for test or help	3.8	1.5	6.0	1.5	.05
Off-task (sitting)	8.8	2.8	5.7	2.3	.10
Off-task (moving in class)	2.0	1.3	0.7	0.3	.05

* Four slow students and four fast students were observed. Each student was observed five times during the year for a period of 20 minutes. Total observation time on each student was 100 minutes.

the same amount of time arranging assignments. The number of waiting occurrences was significantly higher for fast students.

As noted above, the fast student worked more and idled less than the slow one. If our first assumption that curriculum adaptability can be detected from measures of on-task behavior is correct, this result indicates that the present curriculum is not equally adaptive to the entire range of individual differences. The term curriculum is used here in its broadest sense. It is not just the instructional unit or a sequence of units, but also the kind of decision made by the teacher about presenting a specific curriculum unit to a particular student. If it can be assumed that the teacher generally made the right decisions within the context of what could possibly be done, curriculum developers should then recognize the limits of the present curriculum to attract and hold the attention of different students.

In any case, it seems that it is not the theory of the "proper match" but rather the actual present attempt to approximate that match with a wide range of individual differences which needs to be reconsidered. Several reasons can account for this situation: (1) some students (slow ones) require more gradual sequencing; (2) some prerequisites for achieving the tasks may not be previously obtained; (3) differences in interests are not met by the present curriculum; and (4) available curriculum units are not varied enough to equally attract all students.

The fact that both the slow and the fast students spent about the same percentage of time in arranging assignments may indicate that this amount (about 14 percent) is a characteristic of the IPI procedure rather than a difference between slow and fast students. Interestingly, with regard to this category, the standard deviation among the slow students was three times higher than that among the fast ones. It should be remembered that slow students have, in fact, less cause for spending

time in arranging assignments since they master fewer units in a period of time. Therefore, it seems that being busy with arrangements may have more than one purpose for some of the slow students. Perhaps making arrangements becomes an end in itself or, perhaps, some of the slow students are hampered by the need to arrange their assignments.

With regard to "waiting" time, the difference found between slow and fast students was in the number of occurrences of waiting behavior (but not the total amount of time spent waiting). This difference is not surprising if one remembers that fast students master more units, and that any progress to a new unit must be approved by the teacher. Moreover, the fast students not only master more units, they also face more points of novelty and difficulty; they probably are less reluctant to request help from the teacher, and are more oriented toward academic achievement. It thus seems that one should expect the fast students to exhibit more waiting behavior.

Student-Teacher Interactions. The frequency of interactions for giving general attention was less than 1 p[illegible]nt in the behavior of both slow and fast students, and thus is not included in the tables. The infrequent occurrence of this behavior may indicate that disciplinary problems were uncommon in this particular classroom. This finding is in accordance with the general belief that in individualized instruction, in which every pupil is free to work at his own pace, there is a minimal need to control or to discipline students.

The number of occurrences of student-teacher interactions which concerned on-task activity was significantly higher with regard to the fast students compared to the slow ones. (See Table 3.) This finding is apparently a result of the previously mentioned higher rate of "waiting" frequencies among fast students. If this analysis is correct, it can be

Table 3

Mean Percentage of Time and Number of Occurrences of Student/Teacher Interactions

Activity	Slow Students*		Fast Students*		t-test
	Mean	S.D.	Mean	S.D.	(p. level)
Guided by the teacher (total amount of time)	8.8	1.9	9.2	3.7	N.S.
Guided by the teacher (number of occurrences)	2.9	0.4	3.6	0.3	.025

* Four slow students and four fast students were observed. Each student was observed five times during the year for a period of 20 minutes. Total observation time on each student was 100 minutes.

said that teacher behavior in this classroom is in accordance with the principles of adaptive education, i.e., the teacher is responsive to the apparent needs for help. There are, however, two points of concern. First, if teacher behavior is a simple function of students' requests, the teacher, then, does not initiate interactions with slow students who may otherwise not request help. There is some evidence (Klein, Roden, Gentile, Resnick, Reynolds, & Bachmeyer, 1972) that increasing the number of interactions between teachers and slow students results in higher on-task behavior. These interactions probably have an important reward effect on the student.

Second, as might be otherwise predicted, there is no evidence from the data gathered in this study of teachers favoring particular kinds of students. The fact that fast students had significantly higher frequencies of teacher guidance can be explained, as indicated above, by teachers' responsiveness to the greater number of requests for help from the fast students. In the psychological literature, some psychologists have asserted that favoritism does exist. For example, Rosenthal and Jacobson (1968) and Good and Brophy (1969) claimed that teachers' expectations of student performance function as self-fulfilling prophecies, such that positive teacher expectations tend to increase student performance. Good and Brophy found in the traditional classroom "pro-active teacher behavior that goes beyond the objective differences among the children and suggest that teachers may be enhancing these differences rather than reducing them through compensation technique." Again, even though this study did not concentrate on this particular question, the evidence gathered does not substantiate a claim of teachers' favoritism.

Student-Student Interactions. The remarkable finding in this domain is that occurrences were recorded in the "dialogue with neighbor"

category in the off-task dimension (see Table 4), but almost no peer activity was recorded in the on-task categories ("help other student" and "guided by student"). Apparently, the IPI system has not yet made provisions (at least in this particular classroom) for peer learning and peer tutoring whereby students can help each other or cooperate in a common learning task.

Nevertheless, both slow and fast students spent about the same amount of time in dialogues of the off-task kind (between 10 to 15 percent of their time). This is not a negligible amount of time and might indicate a real need for social interaction, a need that does not change drastically even among academically motivated students. It might be helpful, therefore, to channel some of these social interactions into the learning on-task domains.

Group Activities. The degree of social activity (see Table 5) which involved more than simple dialogue was very low for both fast and slow students. The total percentage of time spent in this activity was about 1 percent. On the average, the number of interactions that occurred in 20 minutes of observation was less than one, and there were no significant differences between slow and fast students in this respect. There are two possible explanations. First, more social activity goes on in the classroom in the afternoon than in the morning when the observations were made. Second, math and reading are the only subjects studied totally on an individual basis.

Summary and Educational Implications

It appears that the observational system employed in this pilot study is sensitive to individual differences since the results clearly show two distinct patterns of behavior for slow and fast students. These

Table 4

Mean Percentage of Time Spent in Peer Interactions

Activity	Slow Students*		Fast Students*		t test
	Mean	S.D	Mean	S.D.	(p. level)
Dialogue with neighbor (off-task)	15.5	7.8	9.2	5.1	N.S.
Social activity (off-task)	1.2	1.5	0.9	0.9	N.S.

* Four slow students and four fast students were observed. Each student was observed five times during the year for a period of 20 minutes. Total observation time on each student was 100 minutes.

Table 5

Mean Number of Occurrences of Peer Interactions

Activity	Slow Students*		Fast Students*		t-test
	Mean	S.D.	Mean	S.D.	(p. level)
Dialogue with neighbor (off-task)	5.9	2.2	5.4	1.2	N.S.
Social activity (off-task)	0.9	0.9	0.7	0.6	N.S.

* Four slow students and four fast students were observed. Each student was observed five times during the year for a period of 20 minutes. Total observation time on each student was 100 minutes.

findings permit discussion about the way the educational system adapts itself to some aspects of individual differences.

It is often asked how to explain the slow progress of some students: Is it because the slow student works slowly or because he simply works less? A major finding of this study clearly indicates that much of the variance in the behavior of slow and fast students can be attributed to the fact that the slow students spend less of their time in on-task activities. If this finding is universally true, it should be faced by curriculum developers, instructional designers, and all others for whom adaptive education is a major concern. The display of curriculum of adaptive education (see Resnick, 1972) should vary in such a way that all kinds of students find it highly attractive.

In the classroom observed in this study, much of the control of the teachers' time was left to the students themselves. In such a situation, the fast students will always interact more often with the teacher. If adaptive education were to become compensatory for those who come to the school with obvious disadvantages, one would expect teachers to plan their activities in such a way that interactions with slow students will not depend on student initiative alone. Perhaps one way of achieving this would be to provide the type of data gathered in this study to teachers so that they can plan the distribution of their time in a manner which is more sensitive to individual needs.

Since the findings of this study showed that the fast student worked longer, idled less, and received the teachers' help more frequently, one may conclude that the system favors the fast student. Nonetheless, it should be remembered that the production of the fast student, in terms of the number of units mastered, is found on the average to be three times higher than that of the slow student. Hence, if the total amount of

time spent in each category is divided by the number of units mastered during the period of observation, every difference found in the study will simply change its direction. That is, when total amount of time spent in each category is divided by the number of units mastered in this same time, the fast student appears to spend relatively less time in on-task behavior in terms of time per unit, and has the teacher's guidance less frequently. What emerges from this analysis is that in terms of educational "cost," the slow student is much more costly to the system. Consequently, it might be said that the IPI system (as well as many others) invests more in the education of the slow student than the fast one.

And, finally, the IPI system, as many other early education projects, has a generally high adult-to-student ratio. Yet, even this high ratio cannot satisfy the students' demand for teacher guidance. Both slow and fast students spend about 10 percent of their time waiting for the teacher. They spend more time waiting than being guided and the number of times they indicate waiting is persistently higher than the number of times they receive the teacher's attention. If there is no way to increase the adult-to-student ratio, maybe alternatives ought to be considered, one of which is for the students to help each other. Another way to solve the guiding problem is to increase self-controlled tasks (Wang, 1973) which presumably require less teacher intervention for some of the students.

References

Bolvin, J. O., & Glaser, R. Individualized instruction. In D. W. Allen & E. Seifman (Eds.), The teacher's handbook. Glenview, Ill.: Scott, Foresman, 1971. Pp. 270-279.

Cronbach, L. J. How can instruction be adapted to individual differences? In R. M. Gagné (Ed.), Learning and individual differences. Columbus, O.: Charles E. Merrill, 1967. Pp. 23-39.

Glaser, R. Adapting the elementary school to individual performance. In Proceedings of the Invitational Conference on Testing Problems. Princeton, N. J.: Educational Testing Service, 1968. Pp. 3-36.

Glaser, R. Individuals and learning: The new aptitudes. Educational Researcher, 1972, 1(6), 5-13.

Good, T. L., & Brophy, J. E. Analyzing classroom interaction. Austin, Tex.: University of Texas, 1969. Reprint Series No. 26.

Klein, R. D., Roden, A. H., Gentile, J. R., Resnick, L. B., Reynolds, L. J., & Bachmeyer, B. The effects of a systematic manipulation of contingencies upon overt work behavior in a primary classroom. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1972. Publication 1972/16.

Lindvall, C. M., & Bolvin, J. O. The project for individually prescribed instruction: The Oakleaf project. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1966. Working Paper 8.

Lindvall, C. M., & Bolvin, J. O. Programmed instruction in the schools: An application of programming principles in "Individually Prescribed Instruction." Sixty-sixth Yearbook of the National

Society for the Study of Education, Part II. Chicago: NSSE, 1967. Pp. 217-254.

Resnick, L. B. Design of an early learning curriculum. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1967. Working Paper 16.

Resnick, L. B. Open education: Some tasks for technology. Educational Technology. 1972, 12(1), 70-76.

Rosenthal, R., & Jacobson, L. Pygmalion in the classroom. New York: Holt, Rinehart, & Winston, 1968.

Wang, M. C., Resnick, L. B., & Schuetz, P. A. PEP in the Frick Elementary School: Interim evaluation report of the Primary Education Project, 1968-1969. Pittsburgh: University of Pittsburgh, Learning Research and Development Center, 1970. Working Paper 57.

A DEVELOPMENTAL STUDY OF GROUP FORMATION IN CHILDREN[1]

Donald M. Omark and Murray S. Edelman

Committee on Human Development
University of Chicago

When they start school children frequently confront large numbers of peers for the first time. During a relatively short period of time children establish a position for themselves within their classroom, make friends, and learn to cooperate with others. An ethological approach to the analysis of this process would suggest that the ways in which children begin integrating themselves into classroom social structures should show phylogenetic similarities with the social interactions performed by the juveniles of other primate species. These similarities could occur on the levels of: (1) gestures expressed toward peers, (2) the amount and type of physical interaction, (3) spacing patterns and activity levels, and (4) the total organization of group structures.

In traditional ethological studies, the research procedures would include making detailed records of particular behavior patterns. These patterns would be described in terms of their physical attributes without any motivational causes being ascribed to them. The total set of patterns

---

[1] The results reported in this study were supported in part by the United States Office of Education and the National Institute of Education The opinions expressed are not to be taken as reflecting the view of either the Office of Education or the National Institute of Education.

or behavioral characteristics would constitute an "ethogram" of the species being studied. A variety of individuals within a species would be investigated, e.g., young, adult, male, female, and the portion of the ethogram performed by the various individuals would be noted. Sequences of action and response between individuals would also be investigated and then used to generate hypotheses about the relationships between actions and between the individuals performing them (cf. Blurton Jones, 1972). The basic methodology within this approach is naturalistic observation and the fineness of detail is dependent on the particular recording devices used, e.g., observing, filming, videotaping, etc.

The behavioral characteristics that emerge from ethological studies would be expected to constitute behavioral elements that, to some larger or smaller extent, are: (1) universal to the species, (2) exhibit sex differences, and (3) exhibit individual differences resulting from particular genetic constitutions and/or particular environmental experiences. Behavioral repertoires of species should contain some elements of behavior that exhibit phylogenetic continuities, and some elements that are species specific. The first reflects adaptations that were necessary for survival across species encountering broad similarities in environmental situtations; while the second reflects adaptations to the species historically specific environment. For example, vocalizations such as crying might be expected to occur throughout the mammalian phylum as a signal indicating that severe environmental pressures are being experienced by the infant, and that he requires the mother's attention. But crying as a component of loneliness would not be expected to occur in those species which experience regular predator pressures and where the infant may be left alone for extended periods of time (e.g., rabbits, deer, etc.). Thus selective pressures will elicit or suppress

behaviors in species experiencing different environmental situations. Learned or conditioned behaviors are then modifications of this basic repertoire, but these behaviors are always limited by the possible set of behaviors that the organism is capable of performing.

The same ethological methods and perspectives are used in the study of human behavior. From the literature on primates, the human ethologist obtains clues about what to look for. In the present study, the approach of proceeding from data to determination of patterns to generation of hypotheses was reversed because of the extensive primate literature which exists. This literature indicates that the process of forming dominance hierarchies among peer groups may also occur in children. Children's concept of the social organization of their world was determined by asking the children about it. First the children's free play activities were observed to determine if their group str . .s were similiar to those found in primates and, hence, could lead to hierarchical social structures which might be presumed to resemble those found in primate troops.

Rationale

A basic feature of social structures in primate troops is the dominance hierarchy. This hierarchy is a linear or quasi-linear arrangement of individuals that provides differential access to portions of the environment. It may also provide a locus of attention within which certain individuals can function as troop leaders, and it definitely provides a group structure which minimizes inter-individual conflict. One of the problems facing the young as they grow and develop is to be able to establish and maintain a position within this hierarchy.

A similar problem appears to face human children. For eventual functioning within adult society they would need to first be able to establish and maintain contact with their peers. For each child this involves being able to recognize relationships among peers, and making a place for himself within these interaction networks. Of perhaps greatest importance in this process of moving toward others is for a child to be able to maintain a high feeling of self-esteem until his position within his social world is fairly secure.

In this study, phylogenetic continuities were sought between patterns of social interaction in children and other primates. The interactions examined were then presumed to affect the children's perceptions of their peers. In particular, the children were questioned to find out if dominance hierarchies occurred within their classes, and what the developmental changes were in these hierarchical structures. Four hundred fifty children from a private middle-class school were observed and tested. They ranged in age from nursery school through third grade. A variety of methods were used but because of time limitations only three will be described: (1) group size, (2) nearest neighbor data (collected on the playground), and (3) a hierarchy test (given in the classroom).

Playground Observations

A total of 435 children, enrolled in preschool through grade three of a single school, served as subjects for the present study. Two different observational methods were developed for the study, and the two observers employed for the study were specially trained to use these observation methods.

The first method--group size--involved watching the playground for a few minutes to see which children were moving together in games

or other activities, and then scanning the playground from right to left recording all of the groups as they occurred. Children moving together, e.g., as in a game of dodge ball, were counted as being in a group. Children stationary, but within one meter of another child, were counted as being together, even if they were not interacting. A child standing alone was counted as a group of size one.

The second method--nearest neighbor--recorded individual spacing patterns and modes of interaction (Kummer, 1968; Sommer, 1969). A scoring sheet was used and is included in Appendix E of this paper (see Criteria for Observations). In particular, the four categories of behavior that were recorded on the sheets were: (1) the first, second, and third neighbor of the child being observed, (2) the relative distance between the child and each of his neighbors, (3) the type of interaction, if any, occurring between the child and his neighbors, and (4) the activity level of the child during a portion of the observation period. A first neighbor is defined as the child spatially closest to the child being observed, the second neighbor as second closest, etc.

The observation period for each class was the morning recess period, 15 to 20 minutes in length. The classes tended to stagger their recess periods so that normally an entire class could be observed without confusion. When the children were of more than one grade level the size differential between the children permitted recording to continue. In the few instances when more than one class of the same level were on the playground the children were recorded as being in kindergarten, first grade, etc., and not in a particular class.

To be sure of observing each child, the number of boys and girls in a class were noted as they came onto the playground. The child to be observed was selected by dividing the playground into two halves.

A boy and a girl were observed on one half, and then a girl and a boy were observed on the other half. A short description was made of each child so that he or she would not be observed twice. The child was selected, watched for approximately ten seconds to familiarize the observer with the situation, and then the child's interactions, distance from neighbors, etc., were recorded for the next 15-second period. The distance covered--the activity level of the child--was recorded during the subsequent 30-second period.

This present study was an attempt to relate children's grouping tendencies to the actions performed by other ground-dwelling primates. The types of interaction which occurred on the playground are grouped categories of behavior rather than descriptions of finely detailed physical movements as found in an ethogram. There are two reasons for this procedure. First, the broad patterns, together with the children's perceived hierarchies, would be sufficient to generate hypotheses about both children's and primates' group behavior that could be examined in more detail later. Second, an ethogram is meant to be a nearly exhaustive portion of an organism's behavioral repertoire. Hence, the preparation of one is fairly time-consuming. Since the rules of data collection require that the definitions of the behavior patterns be complete before data collection begins,[2] the present study served as a time for preparing the

---

[2]The logic behind this "rule" is that the apparent first appearance of a new behavior pattern in the middle of a study might lead one to want to begin recording its occurrence, use it for showing differences between groups, etc. In fact, it might have appeared earlier but it simply took some time before the investigator became aware of it. Hence, its sudden inclusion might weigh against a group where it had occurred but had not been noticed. Scoring from filmed records obviates this problem.

ethogram while collecting other data that could be immediately analyzed. The ethogram was then used on a subsequent study. The categories of behavior listed under "Criteria for Observations" (see Appendix E) were chosen to be broad enough to give a feeling of what occurs on the playground while still indicating areas where sex and age differences might be profitably explored with the finished ethogram. These later results are now in preparation. (For human ethograms already completed, the reader is referred to Blurton Jones, 1972, and McGrew, 1972.)

Results of the playground observations. In the grades after nursery school, the boys' groups were larger than the girls' groups (Table 1). This was both in terms of the maximum and average membership of the groups ($p \leq .05$). The average size of the groups increased for both sexes as they matured, but the boys' rate of increase was higher. The maximum size of the boys' groups continued to increase across the grades, while the girls appeared to have a maximum of five or six children at any grade level. The very large boys' groups usually contained a few girls, and in those classes where the children were known to the observer those girls tended to be near the top of the girls' hierarchy.

Table 2 shows that boys played with boys, while girls tended to be near girls. Boys were the three nearest neighbors of boys in 48.5 to 74.0 percent of the observations; the girls, similarly, had girls as neighbors in 48.5 to 69.9 percent of the observations across the grades from nursery school to second grade. Girls were also near teachers for a significantly larger percent of the observations than were the boys ($p \leq .005$). For both sexes, contact with an adult declined steadily with increasing age; the largest decrease occurred after nursery school where over 20 percent of the observations showed both sexes near a teacher. The high percent of time nursery school children were near teachers may be due in part to their also being observed inside during "free time."

Table 1

Maximum and Average Number of Children in Groups of Predominantly One Sex (> 60 percent)

Predominant sex of the groups		Grade			
		N	K	1	2
Boys	Maximum	6.00	10.00	.11	16.00
	Average	3.36	2.28	3.46	4.55
Number of groups observed		40.00	200.00	75.00	18.00
Girls	Maximum	5.00	6.00	5.00	6.00
	Average	3.86	1.92	2.16	3.60
Number of groups observed		29.00	163.00	118.00	20.00

Table 2

Percent of the Three Nearest Neighbors of the Observed Child Who Are Boys, Girls, or Adults (n^1=number of observations on children)

Child being observed		Grade			
		N	K	1	2
Boys	Boys	48.5	65.2	74.0	66.3
	Girls	30.4	28.3	21.9	28.5
	Adults	20.1	6.5	4.1	5.2
	n	105.0	211.0	181.0	84.0
Girls	Boys	29.1	40.2	19.6	39.0
	Girls	48.8	48.5	69.9	52.4
	Adults	22.1	11.3	10.5	8.6
	n	95.0	195.0	217.0	86.0

Physical interaction, such as playful wrestling, holding hands, or throwing a ball to one another, increases with age for both sexes, but it reaches a peak for boys earlier than for girls (Table 3). Agressive and verbal encounters were recorded separately. Boys were found to be significantly more aggressive than girls across the grades ($p \leq .009$) with peaks in kindergarten and first grade, while girls were more verbal than boys at all grades ($p \leq .005$).

For all grades from nursery school through second grade the boys were more active ($p \leq .05$). Boys also maintained a greater distance between themselves and other children ($p \leq .05$), and between themselves and any teachers present on the playground. If opposite sex neighbors happened to occur they were significantly farther apart than same sex neighbors ($p \leq .001$).

In general, the playground groups of young children were found to be like the social interactions which occur among ground-dwelling primates. The boys gathered in larger and more active groups than did the girls. Boys played with boys, while girls played with girls. The girls also associated with adult females more frequently than did the boys. The boys were more aggressive but, as with some juvenile female primates, some of the girls could be found in the rough-and-tumble boys' groups (Kummer, 1968).

Hierarchy Test

The existence of dominance hierarchies in primate troops is generally inferred from the physical encounters which occur between troop members. The recording of aggressive encounters on the playground only indicated that boys fought more than girls. A hierarchy could not be derived from these observations because the individual children

Table 3

Percent of Children Observed, of Each Sex, Who Were Physically Interacting with Their Nearest Neighbors on the Playground (n as in Table 2).

	N	K	1	2
Boys	29.6	32.7	96.1[1]	75.0[1]
Girls	23.1	19.4	27.2	71.8[1]

[1] An interaction with one neighbor counts for one encounter, with two neighbors counts for two encounters, hence, the high percentages mean that some of the children are interacting with more than one neighbor.

involved were not recorded. Instead of making more detailed observations, we questioned the children about the dominance hierarchy which they perceived within their classrooms. The basic question asked was, "Who is the toughest?"[3] Two different forms of the test were used depending on the age of the children.

In nursery school and kindergarten the children were photographed and snapshot size pictures were made of each child. On the day of the test each child was individually taken out into the hall. He was shown the photographs of his classmates, placed horizontally on a bench or on the floor. The photographs were arranged in alphabetical order by first name.

The instructions were: "I'm going to ask you some questions about your classmates. The first question is about toughness. Now what is another word for tough?" (If the child had trouble answering, he was told, "Do something tough.") Now let us look at the first child in the row. If that child is tougher than you, turn his picture over." After the child acted, the experimenter made sure he understood the question. The experimenter then repeated the question with each picture until he was confident that the child understood the task. Then he said, "Now continue on down the row, turning over the picture of each child that is tougher than you."

From first grade through third grade a paper and pencil test was used. The children's names were listed alphabetically down the side of a sheet of paper. The paper was cut so that tabs with each child's

---

[3]Other questions were: "Who is the smartest, nicest, and has the most friends?" Of these, "toughest" had the earliest and highest dyadic agreement. Research is continuing in this area.

name could be easily torn off. The children were also given a sheet of paper with the numbers one through the size of the class listed down the left-hand side. The children were told to "Look through the list of names and tear off the name of the child who is the very toughest in the class. Place that name beside the number one. Now look through the list and find the next toughest child and place that name beside the number two," etc.

Results of the hierarchy tests. The distribution of children across a dominance hierarchy results in boys being placed near the top, girls near the bottom, and considerable overlap in the middle (Table 4). This is the same pattern found in many primate troops (Carpenter, 1964). As can be seen in Table 4, the configurations of the hierarchies change with age.

The basic unit of analysis for the hierarchy test was the dyad of established dominance. The percent of dyadic agreement was formed by taking those pairs where both children agreed on who was the dominant child and dividing by the possible dyads from the class. For example, if John said that Bill was "tougher" than John, and Bill said that Bill was "tougher" than John, this pair was said to be an established dyad on "toughness."

Following the ages suggested by Piaget, it was hypothesized that the school age child (age 7, first grade in our sample) could readily perceive a hierarchical relationship and, hence, would have a higher dyadic agreement than the preschool child. This was tested on the kindergarten through third grade sample. In examining all of the possible dyads within each class it was found that there was a highly significant linear trend ($F = 79.2$, $p \leq .0001$). The largest jump in agreement on relative status occurred between kindergarten (40 percent) and first

Table 4

Dominance Hierarchy Distribution of Children by Quintile Rank at Each Grade Level (Percent of Each Sex).

Grade	Sex	Quintile Rank					
		N	1	2	3	4	5
N	Boys	22	27	23	23	23	5
	Girls	19	5	5	26	26	37
K	B	69	29	35	25	10	1
	G	47	....	2	13	46	40
1	B	54	31	35	22	7	4
	G	50	4	6	12	36	42
2	B	55	25	38	25	5	5
	G	45	4	....	13	33	49
3	B	38	32	42	16	8	3
	G	36	....	....	25	33	42

grade (62 percent). This large increase in percent of dyadic agreement corresponded to a similar increase on a smaller sample where the underlying cognitive level was measured (see Edelman, 1973). Children seemed to develop a consistent perception of their dominance structure at the same time as they were developing the logical operations of seriation and transitivity.

The peak in aggressive encounters on the playground for the boys extended across kindergarten to first grade, and the children's perceptions of dominance relationships appeared to follow from this experience. As discussed elsewhere, the younger nursery school children fought much less and were almost unable to form a hierarchy (Edelman & Omark, 1973a). They gave extremely self-centered responses, almost always placing themselves first or second in any hierarchy, but experiences in kindergarten and first grade seemed to structure their perceptions in a way that substantiated the view that a dominance hierarchy existed within each classroom.

From the results of the primate studies, it was hypothesized that most agreement on "toughest" would be boy-girl pairs followed by boy-boy pairs and then girl-girl pairs. This hypothesis was clearly confirmed by the results. Boy-girl pairs had an agreement of 69.4 percent, boy-boy pairs had an agreement of 54.9 percent, and girl-girl pairs had 51.7 percent. The difference between cross-sex dyads and boy dyads was significant ($F = 16.5$, $p \leq .0008$), as well as the difference between boy-boy and girl-girl dyads ($F = 5.02$, $p \leq .036$).

Accuracy

On the playground, boys were found to play with boys, and girls played with girls. The boys were also in larger groups and more

frequently involved in rough-and-tumble play, while the girls quietly talked in groups of twos and threes. Hence, the boys were seen as tougher than the girls, and they appeared to have worked out the status relationships among each other more completely. Considering the amount of involvement within same sex groups, it was rather surprising to find that each sex could produce the same hierarchy for the opposite sex as that sex produced for itself.

The children were scored for their accuracy of perception. The percent accuracy of perception measured each child's perception of dominance in those pairs of established dominance of which he or she was not a member (number of correct choices of the dominant child in the dyad of established dominance divided by the total number of dyads where agreement occurred). Although there were differences in dyadic agreement among the different sex pairings there was no difference in the percent accuracy of perception for these sex pairings ($F = 1.17$).

Not only did boys and girls have a similar level of agreement about their own sex group, but the rank orders produced by each of the sex groups were highly correlated. The average correlation in a class between the average rank orders produced by the boys and the order produced by the girls when ranking the boys was 0.86 and when both groups ranked the girls the average correlation was 0.79.

Therefore, although the males were more involved in working out their dominance relationships with each other and, hence, have a more clearly defined dominance order, the girls can perceive the dominance relations of both boys and girls as accurately as can the boys. This finding supports the parallel to primate social structure suggested by Chance and Jolly (1970) that stable group functioning is dependent upon all members of the group paying attention to the dominant members.

Overrating

Entrance into a dominance hierarchy would not appear to be an easy task. As seen earlier, with children, or at least boys, this means involvement in aggressive encounters and, at best, numerous knocks to one's ego. The egocentricity of nursery school children suggests a means through which this entrance may prove to be less than traumatic. As found by McGrew (1972) in his study of preschoolers, it did not seem to matter in terms of children's subsequent behavior whether they lost an encounter to another child of lower status than themselves. A similar result seemed to occur, at least for the boys, in this study.

Children of all ages agree on their relative status relationships with many of their peers, and their agreement increases with age--from virtually no agreement beyond the boy-girl difference in nursery school to 66 percent agreement in third grade. Of the remaining pairs, both children could say that the other child is tougher, or each of them could say that the self was tougher. This latter was termed overrating and a percent of overrating was formed by taking all of those pairs and dividing by all of the pairs in which disagreement occurred. Table 5 shows the results. Boys were found to overrate themselves compared to other boys significantly more than girls overrated themselves compared to other girls ($p \leq .006$). Eighty to 90 percent of the boys' pairs where disagreement occurred overrated themselves from kindergarten through third grade. In contrast, the girls' amount of overrating changed markedly during this period (K = 98 percent, first = 28 percent, second = 59 percent, and third = 81 percent).

For both sexes, the high amount of overrating and low rate of agreement between pairs in the preschool years reflects the children's egocentrism. While becoming more accurate in their perceptions of

Table 5

Analysis of Variance of Grade and Sex Effects for Percent of Overrating for the Dimension, "Toughest"

	F	p. ≤
Grade effects		
Linear	7.3	0.01
Quadratic	52.7	0.0001
Remaining effects	6.4	0.02
Sex Effects		
Male Dyads versus Female Dyads	9.9	0.006
Cross-Sex versus Male Dyads	2.16	0.16
Grade by sex interaction	6.6	0.0009

self and others as they grow older, the boys maintain an element of egocentrism. If there is doubt about a relationship, it is decided in favor of the self. The boys confront their peers and if there indeed is a dominance hierarchy, then this overrating would appear to be an important perceptual characteristic for them because many losses may not matter if the other is seen as being lower in the hierarchy.

Girls engage only a few others at a time and their hierarchy is not as well defined. Their shift from overrating to underrating in first grade may be their response to adult authority, i.e., to be "good," as well as a withdrawal from the rough-and-tumble world of the boys--a time for watching and learning about others.

From second to third grade, both sexes also exhibit tendencies to overrate. Although this increase was not previously hypothesized, some basis for explaining the increase can be found in the work of Piaget. In his study of moral judgment, Piaget (1965) differentiated between two attitudes toward following rules. At ages six and seven, the child regards the rules of the game as fixed and eternal. At a later age, the child believes that rules are more flexible and that he may influence how they might get changed. In the case of making hierarchies, the first and second grade children were trying to report the "correct hierarchy," but in third grade the child may have been aware that he or she was making the hierarchy and could put people anywhere he or she might wish. When pilot testing the hierarchy test, we found that the younger children asked if they were "right" after they formed the hierarchy, but the older children did not ask if they were right. Sometimes the third graders would put themselves high in the hierarchy and smile impishly, as though they just wanted to see themselves high.

Conclusions

Some of the social actions of children fit nicely within a primate paradigm. This is most evident with sex differences in the size of the children's groups, their amount of association with an adult, and the levels of activity and aggressive actions. The children's perceptions of dominance hierarchies within their classrooms resemble the hierarchies derived from the actions of primates.

The boys were found to have more agreement on their portion of the hierarchy than the girls, but that boys are tougher than girls had even significantly more agreement for both sexes. In spite of the apparent lack of contact between the sexes during play periods, the children's direction of attention included the opposite sex. It was found that each sex could accurately perceive the hierarchy created by the opposite sex. Boys were more involved in aggressive encounters; an apparently necessary corollary of this was that the boys overrated their own status position. Girls were not as involved in forming hierarchies and showed a period during which they underrated themselves.

We are continuing research in the area of the relationship between cognitive development and children's social experiences. Both the ability to seriate and to perceive transitive relations would appear to receive reinforcement from social encounters. If so, then the real world of social interactions with peers might be a very necessary part of each child's educational experience.

From an ethological perspective the interplay between cognitive development and social experiencing would be expected to have phylogenetic as well as individually historic components. Humans and other primates must interact in a social world so that one would expect adaptations for adequate survival in relation to that world to occur through

natural selection. In essence, this means that throughout a child's development, including his cognitive development, he would be expected to exhibit characteristic behavior and thought patterns that enable him to function adequately in the social world with which he interacts at any particular age.

The way children conceptualize their world at each age level can then be seen not only in how they manipulate sticks, volumes, or masses, but also in the way they interact with and perceive their social world. The amount of agreement about one aspect of the children's peer groups that was found at a very young age indicates that the child's ability to seriate appears as early for social interactions as it does for the physical world. Another portion of the study indicates that this also occurs for transitive relations, and that a great deal of emotional involvement is attached to perceptions of social interactions (Edelman & Omark, 1973b). Agreement about the structure of the social world occurred where direct interaction was very limited, as in the low level of interaction found between boys and girls on the playground. Children's abilities to pay attention to others is viewed as a very necessary adaptation that enables them to organize and develop their cognitive abilities. The age changes that occur in these abilities are seen as adaptations that enable the children to continue to move toward peers without being swamped by information and experiences that they are not yet able to assimilate.

If ethology has a message that is relevant for child development, it is that a broader perspective needs to be taken when viewing development. This perspective must include what the child brings to each encounter with his environment from his phylogenetic past. The question always posed by a researcher from the ethological approach is:

"Why does this child act the way he does?" We find it necessary to investigate more than immediate cause-and-effect relationships to find answers to this question.

References

Blurton Jones, N. (Ed.) Ethological studies of child behavior. Cambridge, England: Cambridge University Press, 1972.

Carpenter, C. R. Naturalistic behavior of non-human primates. University Park, Pa.: Pennsylvania State University Press, 1964.

Chance, M. R. A., & Jolly, C. J. Social groups of monkeys, apes and men. New York: Dutton, 1970.

Edelman, M. S. Peer group formation in young children: Perception. Doctoral dissertation, University of Chicago, 1973.

Edelman, M. S., & Omark, D. R. Dominance hierarchies in young children. Social Science Information, 1973, 12(1). (a)

Edelman, M. S., & Omark, D. R. The development of logical operations in the social world - An ethological approach. Paper presented at the Society for Research in Child Development Meeting, Philadelphia, Pa., 1973. (b)

Kummer, H. Social organization of hamadryas baboons. Chicago: University of Chicago Press, 1968.

McGrew, W. C. An ethological study of children's behavior. London: Academic Press, 1972.

Omark, D. R. Peer group formation in young children: Action. Doctoral dissertation, University of Chicago, 1972.

Piaget, J. The moral judgment of the child. New York: The Free Press, 1965.

Sommer, R. Personal space: The behavioral basis of design. Englewood Cliffs, N. J.: Prentice-Hall, 1969.

DISCUSSION

by

Donald M. Medley

School of Education
University of Virginia

The papers presented at this session exemplify a most encouraging trend in research in teaching and learning, and testify both to the health and the potential of that trend. The trend I have in mind may be regarded as an increasing emphasis on means (or process) as a focus of educational research in contrast to ends (or products). We are increasingly interested in "how" rather than "how well" teaching succeeds.

After all, the education enterprise exists solely to maximize the effectiveness of the educational experiences the individual child has in school. We all learned in our first education courses that learning results from the activity of the learner. If this is true, then it would seem obvious that any successful effort to understand the educational process must concentrate on what happens to kids in schools while they are being educated. As one who has always taken this for granted, I find it difficult to understand why so little attention is paid to pupil behavior in so many research studies.

In these five papers, we have five excellent examples of people looking for answers where the answers can be found. I am not surprised to find Omark and Edelman doing this--they come from a discipline that has characteristically studied phenomena where they happen, with a directness that looks almost naive. It would be great for educational research if more of us would take this kind of open-minded look at what children are doing, instead of trying to prove some pet theory of our own, as we usually do.

The rest of our contributors also exemplify what I mean in various ways and varying degrees. Each of them shares an interest in what is happening to pupils. The fact that each has some definite question to ask about what the pupils are doing does not seem to prevent their seeing what is actually going on. This idea of monitoring the implementation of an educational innovation at the point at which it is supposed to take effect (in pupil behavior) is exactly what I am advocating. For half a century, the great majority of attempts to change education for the better that were proposed and attempted were evaluated solely in terms of outcomes. Studies that collected data about whether there was any change in what the pupils did were the rare exceptions. There is no way of assessing the number of promising innovations introduced during these years that were abandoned as ineffective, not because they did not work but because they were never really tried.

Leinhardt's study is almost a textbook example of how a variety of techniques already available to anyone who knows enough to use them can be employed effectively to study the implementation process itself. The study casts light on problems related to the fact that any innovation has different effects in different settings, with different teachers, and with different pupils. Shimron focuses more closely on a single aspect of process that is indeed crucial in the innovation he is studying; and the effort pays off. Wang's paper illustrates the role an ongoing examination of learner behaviors can play in formative evaluation, enabling the innovator to shape the implementation in process. Ross and Zimiles are concerned with the important process of refining an instrument to increase its effectiveness for purposes like these.

These papers share a primary dependence on structured observation instruments, although we have seen, particularly in the work of Wang and that of Leinhardt, useful applications of other techniques.

Structured instruments like the one Ross and Zimiles are developing have unique advantages that usually make them preferable to alternative procedures. The precise, detailed, and objective information they provide about behavior is their principal strength. Even when designed to test specific hypotheses, as was the case in most of these studies, observational data may reveal things not suspected by the researcher because of the quality of the data they yield. In this way, some of the advantages of the openness of the ethologist may accrue--especially with a more carefully refined instrument like the one Ross and Zimiles are working toward.

There are two further developments I would hope to see before too long. To be content with analyzing raw frequencies or percents of certain types of events--however intuitively satisfying it may be--is to stop far short of full exploitation of the data. The data also constitute the raw material for a search for structure in human behavior--for efforts to describe it in terms of fewer and more powerful, meaningful dimensions that are composites of raw frequencies. These are the stuff that theories are made on. If human behavior is indeed lawful, we should constantly be seeking for such laws, and they are to be found in the uniformities, the communalities in the detailed items of our behavior records.

The second development I look for has to do with realizing the basic purpose of research, which is to enable a large number of people to reap the benefits of the researcher's experiences. How effectively this is done depends on how accurately it is possible to communicate just what happened in a study. Structured observational instruments are effective research tools mainly because they facilitate more accurate communication of what happens than alternative procedures do. This

results from the greater clarity the operational definitions of categories must possess--in other words, the greater clarity of the language in which we describe phenomena. Unfortunately, each of us, when he constructs his own instrument, is inventing his own language and vocabulary: and he is, therefore, to a disturbing extent, talking to himself. Every new report requires the reader to learn a new vocabulary. I look forward to the day when we have a small number of standardized observation instruments so that the behavior of pupils observed in many different settings and studies can be described objectively in comparable terms. Then our knowledge may begin to grow in geometric rather than arithmetic increments.

DISCUSSION

by

Lauren B. Resnick

Learning Research and Development Center
University of Pittsburgh

The papers in this symposium use observational techniques for several different purposes in the evaluation and analysis of educational environments. They also concomitantly represent the use of somewhat different methodologies. I would like to begin my discussion with a consideration of the different purposes of the papers in this symposium, and then consider some similarities and differences in their methodologies that seem to be a reflection of these purposes. I will conclude by considering some general questions concerning the role of observational methods in educational research--questions stimulated by the present papers but not specific to them.

Three of the papers, those by Ross and Zimiles, by Leinhardt, and by Wang, are explicitly concerned with some aspect of the evaluation of educational programs. In each case the authors are evaluating their own programs, or those of close colleagues whose theories of education they share. The three studies fall along a continuum that corresponds roughly to the degree of control exercised by the development team over the educational program under study. Leinhardt's paper is concerned with the evaluation of a program as implemented in a wide variety of school settings, settings over which the developers of the program do not have specific control. We might call hers a field evaluation study. This concern with evaluation of the program as it operates under relatively uncontrolled field conditions sets limitations on her methodology,

but at the same time allows a certain breadth of conclusion from her data that is not possible from studies that are made in fewer or more controlled settings.

Ross and Zimiles's work represents a move from broad field evaluation to detailed study of three particular educational environments. They compare their own program with a competing or "control" program; but since they examine two different settings in which their own program is used, they are not concerned strictly with the program's implementation in a given site. Nevertheless, they invest heavily in detailed observation of a few classrooms, as contrasted with Leinhardt's much more generalized observation of a larger number of classrooms. Compared to Leinhardt's approach, Ross and Zimiles's work is more expensive and much more likely to give us detailed information on the way in which children behave in a particular environment, but less likely in its present form to provide an economically viable instrument for large scale evaluation.

Wang's study takes a still further step toward detailed analysis of a specific developer-controlled environment. It is concerned with description and, therefore, evaluation of a specific aspect of a program. The effort depends upon deliberate modification of the program within a single classroom in the course of a school year. Like Ross and Zimiles, Wang studies a broad range of outcomes and, like them, she accepts as a constraint the difficulties of observing in a real setting rather than a contrived or laboratory setting. Nevertheless, Wang introduces an element of experimental control that neither Leinhardt nor Ross and Zimiles attempt. This is accomplished by use of a "reversal design," in which a classroom is first studied under normal or "baseline" conditions; a new program is then instituted, followed by a return to baseline and a return again to the treatment conditions. This strategy avoids the

problems of comparing one classroom to another, with the attendant population variations. Rather it allows comparison of one set of classroom conditions with another set implemented by the same teacher with the same children. Such a study avoids the difficulties of control group comparisons; on the other hand, without extensive replication, it allows us to draw few conclusions about the way in which the special treatment under study--in this case, the self-scheduling system--might work in settings other than the single classroom studied.

Shimron's and Omark and Edelman's studies are more descriptive and explanatory than evaluational in intent. Shimron's study takes on an even smaller range of behavior, in terms of both sample size and variables studied, than Wang's. It focuses on individual differences in the behavior of a small number of children and in the teacher's response to these children, all within a relatively stable classroom environment. Shimron considers the teacher as part of the general environment within which particular kinds of child behaviors take place. He looks for the patterning in these behaviors, and in this respect his work shares some features with Omark and Edelman's. However, having detected behavioral patterns, Shimron seeks to relate them to academic performance; in this respect, he is closest to Leinhardt, who seeks explicitly to measure classroom behaviors that mediate success in academic learning.

Omark and Edelman's paper, unlike the others in this symposium, is not concerned with instruction, but with purely social aspects of behavior in school. Their work draws upon a tradition of research that is only now becoming familiar to psychologists and educators. The ethological tradition of observation, from which Omark and Edelman's work derives, is based in biological studies of animal behavior. As a method for studying children, it brings to the study of human behavior some of the biolo-

gist's concern for detecting patterns of behavior, finding out what children are like in specific environments.

Because of its different substantive focus, Omark and Edelman's work is of primarily methodological interest in the context of this symposium. Their paper does not describe in detail the methodological strategies, but instead focuses on specific theories, drawn from animal observation and now applied to children on the playground and in the classroom. Lying behind these theories, however, is a body of extensive observational technique and literature. As mentioned in their paper, observation typically begins with development of an "ethogram," a detailed description of the sequence of behaviors engaged in by an individual. As repetitive patterns emerge, instruments for systematic observations of particular behaviors are developed. The use of ethological approaches to observation of children's behavior is well described in recent books by Hutt (1970) and Blurton Jones (1972). The important message of this line of investigation for systematic observation of educational processes is the stress on detecting patterns of behavior, on treating early stages of observational work as a time for searching out patterns of behavior rather than for testing formal hypotheses.

The methods of observation used by the authors in these five studies are matched in significant ways to their purposes. Perhaps the most general point to be made in this respect is that whatever method is chosen, there is a "trade-off" of some kind to be made; one gains certain kinds of information at the expense of other kinds. There is no perfect observational system. Observation is expensive and one therefore seeks to maximize the kind of information useful for one's purpose in designing and using an observational system, always recognizing that certain other kinds of information will inevitably be lost.

The result of this is that none of the studies reported here today, and I would venture to say that no single study likely to be produced in the coming y ars, represents a perfect or complete examination of school processes. Yet, each offers something new, either substantively or methodologically, to our understanding of classroom processes. What we can reasonably expect over time is that a combination of different approaches, used for different purposes, in different places and at different times, will gradually allow us to converge on some understanding of what the process of educational interaction is like. My comments on specific studies must be understood in this light. They are offered not so much to criticize the present studies for their lack of particular types of information, but rather as a means of pointing out what each offers and what, in order to offer that, it has had to give up.

A general problem inherent in observational study of classroom behavior is that of cost; this cost needs to be measured in terms of: (a) time needed for the observations to be conducted, (b) reliability both between observers and over time, and (c) difficulty of interpretation of the data gathered. In general, I think it is safe to observe that the more detailed the observations, and the wider the range of phenomena they attempt to describe or account for, the greater the degree to which they encounter such costs. Thus, Wang and Ross and Zimiles encounter greater difficulties in establishing reliabilities and in reducing the complex data they accumulate than does Shimron; but Shimron's high reliabilities and temporally extensive information on individual children depend on limiting both the range of behaviors examined and the number of children observed. Leinhardt, in an effort to keep costs of the various kinds mentioned within reasonable bounds for purposes of field evaluation, uses "index measures" to a large degree rather than direct observations of the behaviors she is interested in. For example, she is

interested in the degree to which assignments are individualized. To study this question in detail would require closely following the behavior of individual children over an extended period of time. She cannot afford this kind of observation and uses instead a count of how many different assignments are given in a particular time period. To that extent the program is being individualized. While the assumption that this measures individualization is not directly tested, it has face validity; and it is indirectly tested later by the extent to which it predicts achievement.

Certain details of the present papers are useful in raising for us some general issues concerning the use and interpretation of observational data. Only a few can be considered in the time available, so I shall comment here on issues that seem to me to be of particular interest. Wang's paper includes in its title, and in its opening paragraph, reference to formative evaluation--that is, the use of observational techniques for providing information to the developer necessary for improvement of a program. Yet the paper itself gives us no detailed examples of how her data were used in modifying the program studied. It would be useful for such descriptions to be made available in the future. Further, now that the self-scheduling program that she describes has been shown to be a reasonably effective one in the classroom, perhaps this suggests that the prescriptive versus exploratory distinction can be dropped and that one should begin to focus on the quality of interaction, regardless of the label given to the activity. Wang's study, in other words, marks a starting point in the study of an educational model; it is in that sense properly labeled "formative." I will be interested to learn the directions in which these investigations lead and to see documented, at least informally, the way in which observational research and program development mutually interact.

In the Wang and Ross and Zimiles studies, ratings rather than direct observations are sometimes used as a way of reducing complex events to a measurable form. It is perhaps appropriate to consider what ratings are and are not useful for in observational work. Ratings are useful, assuming inter-observer reliability is adequate, when one has dimensions already available and wants to know, in a rough way, whether we are approximating our own goals. Thus, they are primarily a device for "summative evaluation," i.e., for deciding to what degree implemented programs meet their goals. Ratings are generally not so useful when one is searching for important dimensions along which behavior varies. They do not allow one to examine relatively "raw" data in search of patterns: to formulate rather than test hypotheses. Given the present paucity of knowledge concerning classroom processes, a case can be made for relying as heavily as possible on direct observation rather than ratings, at least at the early stages of any particular set of investigations.

Ross and Zimiles make no attempt at all to relate the classroom process variables observed to the quality of learning as measured independently of the observational effort. This raises for me the interesting question of whether process in the classroom can be equated with learning outcome. What would happen if children from Ross and Zimiles's traditional public school were suddenly placed in the Bank Street Follow Through classroom? How long would it take for them to behave like the Follow Through children? And conversely, how long would it take for the Follow Through children to behave like the traditional classroom children, if their position were to be reversed. This question is important because it suggests the possibility that the behavior of children may be so heavily influenced by immediate environmental conditions as not to reflect a stable individual pattern at all. Rather, what Ross and

Zimiles may be measuring is the extent to which they and their colleagues have succeeded in creating the kind of environment they want to create. This is an important step in the study of one's educational program, but it is only a step. The next step necessary is to consider whether the program, once created, does indeed have the effects one is seeking. I do not mean to imply that programs should be judged solely on the basis of whether they promote higher achievement scores in mathematics or reading. What I do mean to imply is that some measurement of learning effects that are independent of specific learning environments may need to be sought. This implies a "next stage" of classroom observation, in which individual children are observed in different environments and over extended periods of time. In such a program of research, Ross and Zimiles's work would be viewed as an important first step.

The richness of the data in these studies makes it difficult to do justice in these brief remarks to the research reported. What is perhaps most important to note in conclusion is that this symposium and the papers in it mark a turn toward a new interest in description and analysis of the educational process. This is a "new look" in educational research that, while tentative for the moment in the conclusions it permits, is of potentially major significance for increasing our understanding of and skill in improving the educational enterprise. The five studies represented here, while varying in purpose and method, share a common concern for describing how people--children and teachers--actually behave in school. This is a question about which we now have astonishingly little information, having focused our research over the past several decades very heavily on academic outcomes. It is only by including in our study of education descriptions of educational processes--as outcomes of interest in their own right or as mediators of more traditional outcomes--that we can expect to make sense of the educational

enterprise, to understand how, as well as whether, instructional programs work. This symposium represents a most welcome movement in that direction.

APPENDIX A

Description of the Learning Research and Development Center's Individualized Instructional Programs

by

Margaret C. Wang

Learning Research and Development Center

University of Pittsburgh

APPENDIX A

The LRDC Individualized Instructional Programs

The LRDC Individualized Instructional Programs include the Individually Prescribed Instructional (IPI) Program for children of elementary grades (grade one through six), and the Primary Education Project (PEP) designed for children of early childhood age (age three through seven). Both, IPI and PEP were developed to provide educational experiences that are adaptive to the learning needs of the individual student. The programs were designed with the basic assumptions that: (1) children display a wide range of differences in their entering abilities and the ways in which they learn and acquire competencies, and (2) to provide educational experiences that are adaptive to the individual differences means, to provide learning situations (e.g., classroom organization, learning materials, etc.) that can accommodate the needs of the individual student and when needed, teach the prerequisite abilities demanded by the learning situations (Glaser, 1972).

The LRDC Individualized Instructional Programs are designed with the following guidelines (Glaser, 1970): (1) the goals of learning are specified in terms of observable student performance and the conditions under which this performance is to be manifested, (2) the learner's initial capabilities relevant to for coming instruction are assessed, (3) educational alternatives suited to the student's initial capabilities are presented to him; the student selects or is assigned one of these alternatives, (4) the student's performance is monitored and continuously assessed as he learns, (5) instruction proceeds as a function of the relationship between measures of student performance, available instructional alternatives, and criteria of competence, and (6) as instruction

proceeds, data are generated for monitoring and improving the instructional system.

Curriculum components of the IPI program include the Individualized Mathematics Curriculum (Lindvall & Bolvin, 1967), the Individualized Science Curriculum (Klopfer, 1972), and the New Primary Grade Reading System (Beck, 1972). Curriculum components for PEP include beginning math, classification and communication skills, perceptual skills and the exploratory learning skills (Resnick, Wang, & Rosner, 1973). However, not all of the components listed above are used in the LRDC Follow Through schools.

Aspects of curriculum developed for each of the curricular components include: the specification of curriculum objectives, the sequencing of the objectives, the design of instructional and learning activities and materials, the specification of teacher and student behaviors, and the specification of procedures for diagnosing and monitoring student learning progress. Provision for the diagnosing and monitoring of individual student's learning progress is at the core of the individualized instructional programs. Procedures and instruments (e.g., Cox, 1968; Wang, 1968) for diagnosing and monitoring student learning has been designed to provide teachers with the information necessary for adapting the use of the program components to the individual students, and to communicate, on a substantive basis, with parents and others concerned with the learning progress and the development of the student.

The implementation of the LRDC Individualized Instructional Programs in classroom settings, ideally requires two adults in each class, a teacher and an aide. The adults during the instructional period, generally perform two basic roles, the "traveling" role and the "testing and tutoring" role. The traveling role requires the teacher to circulate

among the students, helping with their learning tasks and checking them off as they are completing and interacting in various ways, generally quite brief. The testing and tutoring role requires the teacher to work intensively with individuals or small groups of students for such purposes as administering diagnostic tests, tutoring individual students, giving group lessons or working with a group of students on a special learning project. The roles described above are "idealized" descriptions; in practice, however, the two adults fluctuate from one role to the other as need arises.

There are two basic sets of teacher functions, both necessary for smooth and effective implementation of the LRDC program in classroom settings, the management functions and the instructional functions. The management functions are concerned with the establishment of an effective system for classroom management. They include such functions as: the provision of materials and equipment for the various components of the program, their physical arrangement, display and storage, maintenance, as well as demonstrating and explaining rules and the use of materials, and praising or otherwise reinforcing students for appropriate self-management activities.

Two sets of teacher instructional functions have been identified: the "didactic" and the "consultant" functions. The didactic instructional functions are related to the administering of tests associated with the formal curricula, prescribing learning tasks on a daily basis, checking prescriptive activities and giving help on them as required. The teacher and/or the aide, also assume, under the didactic instructional functions, the responsibility to conduct special tutoring sessions on certain specified curriculum objectives, as well as large or small group lessons as dictated by the various curricula and by the needs of the students.

The consultant functions are less highly structured, but like the didactic functions, they are carried out in the course of both traveling and testing and tutoring. These functions require the teacher and/or the aide to focus on observation of students' learning processes beyond what is provided in the formal tests; use questioning and probing techniques to stimulate development of self-reflective in problem-solving activities on the part of the students; engage in planning with students, helping them decide what to do and how to do it; pose problems for students to work on and help them in planning and carrying out solutions; and engage in games and other forms of play with the students.

No clear distinction can be made in practice between management and instructional functions--every act contributes to both. Similarly, the teacher should fulfill both didactic and consultant instructional roles. Nevertheless, the distinctions are useful as a means of describing the range of functions that teachers must meet when implementing the LRDC Individualized Instructional Programs. The distinctions between the two functions also serve to characterize, in a general way, the teacher behaviors to be observed in an LRDC classroom.

The role of the student under the LRDC Individualized Instructional Program centers around the management of one's own activities in the learning situations. In general, the student is expected to:

1. Work and complete certain tasks prescribed by the teacher (the nature and the amount varies from student to student, and depends on the learning needs and the individual student characteristics).
2. Work and complete certain tasks of the student's own choice.
3. Make decisions about when to do what work (the range of the options and the degree of control varies from age to age and class to class).
4. Take diagnostic tests when asked by the teacher.
5. Participate in tutoring sessions when asked by the teacher.

6. Participate in group activities when required.
7. Ask the teacher to check the work as one completes the assignments.
8. Ask for help (from the teacher and/or peers) when needed.
9. Assist others (initiates and/or when requested) for management as well as for learning purposes.
10. Follow classroom management rules.
11. Locate learning materials and equipment independently.
12. Carry out material management responsibilities (e.g., clean up, return equipment, etc.).
13. Take turns and share activities and materials with others.
14. Interact with peers for personal as well as school related activities.
15. Tolerate disruption to the activities at hand for attending certain group activities and/or certain testing or tutoring sessions.
16. Attend to the task at hand and ignore distraction from the different activities being carried out by others at the same time.
17. Budget one's own work time to meet the time constraints established for certain tasks.

The student roles listed above are behaviors required to function effectively under the LRDC Individualized Instructional Programs. However, the ability to carry out the roles are not assumed as part of the entering behaviors of all students. Students are taught to acquire the minimum level of competencies required to assume these self-management and independent learning roles.

APPENDIX B

Observational Instruments and Data Recording Forms

Used in

The Differentiated Child Behavior Observational System (DCB)

by

Sylvia Ross and Herbert Zimiles

Bank Street College of Education

APPENDIX B

DCB Form

Bank Street Follow Through DCB (Group) (5/1/71) No. of Children _____ No. of Adults _____

Date _____ Community _____ School _____ Teacher _____

Time from _____ to _____ Actiyity _____ No. of Children _____ Adult Role _____

GIVES INFORMATION RE			ASKS QUESTIONS RE
Identity, Situation			Identity, Situation
Prediction, Plan			Prediction, Plan
Function, Process, Directions			Function, Process, Directions
Differentiating Properties			Differentiating Properties
Relationships			Relationships
Category, Class			Category, Class
Casual Reasoning, Problem Solving			Casual Reasoning, Problem Solving
EXPRESSES			BEHAVES AGGRESSIVELY
Routine Need			Physical - Initiates
Preference, Desire			Physical - Defense
Feelings, Attitudes			Verbal - Inititates
Needs - Task			Verbal - Defense
Needs - Social/Physical			Takes Others Belongings
Affection, Warmth, Humor			Abuses Material -- Equipment
Concern for Others, etc.			Destroys Another's Work
Lack of Concern for Others, etc.			Challenges Established Classroom Limits
SHOWS AUTONOMY*			COMMUNICATES VIA SYMBOLIC PLAY AND REPRESENTATION
Records Choice of Task			Directs Group (Dramatic Play, etc.)
Initiates Task			Projects Fantasy
Initiates Attentional Focus			Makes Descriptive Comparison
Seeks Answers (in Book, etc.)			Narrates Story, Sings Song, etc.
Selects Materials, etc.			Tells Original Story, Plays with Words, Sounds
Replaces Materials, etc.			Shares Reading Experience (Reads Aloud or Shares Pictures)
Straightens Up Work Area			
Resolves Conflict			
Enacts Transitional Move			

* situation - appropriate, no adult intervention

Appendix B (Cont'd)

Bank Street Follow Through

1/18/71

CHILD BEHAVIOR SCANNING FORM*

Community: _____ Observer _____

Grade: _____ Date: _____

Teacher: _____ Time: _____ To: _____

Context: _____

Children	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1. Non-Verbal B																														
1.1 Inv																														
1.2 Sel																														
1.3 Org																														
1.4 Ret																														
1.5 NOF																														
1.6 Soc																														
1.7 Ab																														
2. Verbal B																														
2.1 Cx																														
2.2 Q																														
2.3 RQ																														
2.4 Read																														
2.5 VA																														

List all Activities:

*To be used with DCB form.

1.52

Appendix B (Cont'd)

CHILD BEHAVIOR SCANNING FORM (to be used with DCB)

: Directions and Categories

Directions: A separate entry is to be made for each child in the classroom with the entire group represented on each scanning form.

Start with the child nearest the door then continue in a counter-clockwise direction until the entire group has been scored.

Look at each child long enough to categorize his behavior. It should take from two to five minutes to score the entire group.

Categories.

1. Non-Verbal Behavior

1.1 INV , Involved: task oriented

1.2 SEL Selects materials, equipment, etc.

1.3 ORG Organizes, i.e., arranging materials, furniture, etc; directing others in relation to position in physical setting.

1.4 RET Returns or replaces materials, etc.

1.5 NOF No observable focus, e.g., daydreaming, withdrawn.

1.6 SOC Social interaction (non task-oriented).

1.7 AB Abuses materials, equipment, etc; destroys another's work; physically abuses another.

2. Verbal Behavior

2.1 C X Comments, explains
2.2 Q Questions
2.3 RQ Responds to questions
2.4 READS Reads aloud
2.5 VA Verbally aggressive, i.e. treatens another

Codes. BOY : B (alone). O(with or to adult); △ (with or to another child)
GIRL : G (alone); I (with or to adult); ⚠ (with or to another child)

Context Describe period in general, e.g Discussion, work period

Subject Area. After completing observation list each activity noted.

Community _____ Teacher _____

Observer _____ Date _____

SUMMARY SCALE: CHILD BEHAVIOR (to be used with DCB-Group)

Items to be rated on scale of 1 (lowest rating) to 4 (highest): Circle Number

1 Extent to which children participate with interest and enthusiasm.

1 2 3 4

2. Extent to which children show ability to sustain attentional focus and remain involved in activities.

1 2 3 4

3 Extent to which children remain at tasks until completed.

1 2 3 4

4 Extent to which acting-out disruptive behavior is evidenced.

1 2 3 4

5 Extent to which withdrawn, unrelated behavior is evidenced.

1 2 3 4

Appendix B (Cont'd)

Bank Street College of Education
610 West 112th Street
New York, New York 10025

DCB OBSERVERS

Interview with Classroom Team

Observer: I've enjoyed visiting this classroom very much and would like to ask you a few questions:

1. Did you do anything special today? _____

_____

2. In general, was this a typical day? _____

If not -- how was it different? _____

_____

3. Did the children behave pretty much as they usually do? _____

If not, specify: _____

_____

4. Did each of the children who were observed individually behave as he/she usually does?

_____

If not, specify:
Name of child: _____ Comments: _____

_____

_____

Name of child: _____ Comments: _____

_____

_____

Community _____ School _____ Teacher _____

Observer _____ Date _____

APPENDIX C

Instructions for Observational Techniques

used in the study:

Observation as a Tool for

Evaluation of Implementation

by

Gaea Leinhardt

Learning Research and Development Center

University of Pittsburgh

APPENDIX C

Instructions for Questions 6 and 7

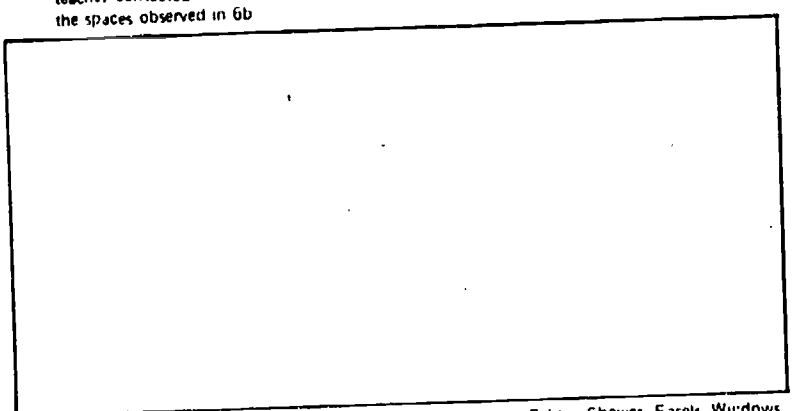
Conditions for filling out 6a, 6b, and 6c: (1) The questionnaire should be filled out during the morning math (IPI or PEP) period, at least five minutes after the children have started their assignments. (2) At least two-thirds of the enrolled children should be present in the room. (3) The recording should not be started if three or more children have already started working in the exploratory area (unless exploratory is always conducted simultaneously with prescription work).

6a. Draw a quick map of the classroom. Divide the exploratory from the prescription and give a general idea in the prescription area of desks or tables, library, shelves, easels, windows, blackboards, audio equipment, TV, and where children are sitting. Also show in a second color the way you divided the room. Then for three minutes, trace the walking path of the teacher with a solid line.

Indicate the verbal contacts (when he/she is not physically present) with a dotted line. If the teacher has traveled rapidly around the room so that your diagram starts to get messy after two minutes, stop and note on the side that you only recorded for two minutes. See example below.

ALLOCATION OF SPACE

6a) Draw a map of the classroom. Draw a pattern of the teacher's contacts for 3 minutes using solid line to indicate the teacher went to the child, and dotted line to indicate teacher contacted child verbally without going to him, in another color lightly outline the spaces observed in 6b



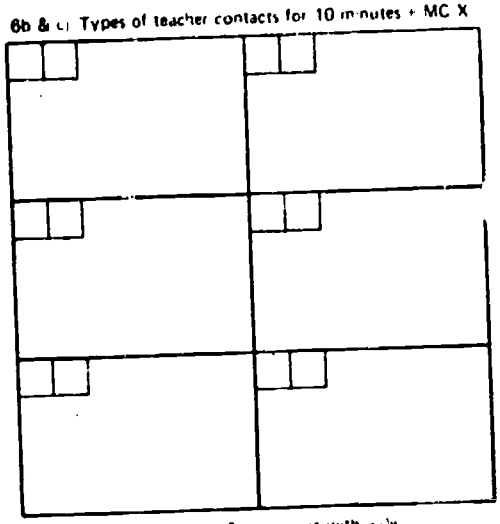
Include Exploratory Area, Prescription Area, Desks or Tables, Shelves, Easels, Windows Black Boards, Audio Equipment, TV, Approximate location of children.

6b. Using the form provided, mentally divide the classroom into six areas using the most natural division possible. That is, divide the room by the way the children and desks are located rather than by equal spaces. Record in each cell whether that cell is a prescription area (P) or exploratory area (E) or both (PE). At the beginning of a ten-minute observation, count how many children are in each cell and record that number in the second upper left hand cell of each of six blocks. Count again at the end of the observation time and record the number beside the first. Children will move around during the 10 minutes of recording; however, the average number per cell should remain relatively consistent. A contact is considered any verbal or physical statement or touch made by the teacher to one or more specific students; it includes general class statements such as: "We are all working well today," or "it's getting a little noisy," etc. Everytime a teacher makes a contact, code it in the block to which the contact was directed, i.e., if the teacher is standing in the upper right-hand corner of the room and speaks to a child in the middle left block of the room, record the contacts in the latter block, not the former. The exception is general contacts; they are recorded in the cell where the teacher stands. The verbal behavior of the teacher should be recorded for 15 consecutive minutes--starting at least five minutes after the children have started math work.

6c. If the teacher is traveling with her aide, use the same way of dividing the room and record the aide's travel for the next 10 minutes in the same way you did for the teacher in 6b; if not, continue with the teacher. (See example below for questions 6b and 6c.)

Appendix C (Cont'd)

6b & c) Types of teacher contacts for 10 minutes + MC X



Is teacher traveling alone ?_____ or with aide_____
Number of adults traveling_____ Time observation started_____

The Codes are:

M Management--A management comment is differentiated from a cognitive one in that it is not subject matter oriented and deals with the overall behavior of the child in the classroom, such as raising or folding hands, sitting at desk, getting box or ticket, etc. In addition, there are management behaviors which relate to the performance of a cognitive task such as getting the materials out, turning to page X, writing the answer on the dotted line, etc. Comments which deal with these types of behaviors are often accompanied by additional or further comments which are cognitive and are coded as such (see example below).

C Cognitive--This is a contact which deals directly with the subject matter content, or the acquisition of subject matter such as sounding out words or counting out chips.

+ Positive verbal praise--Usually this includes an adjective such as good, wonderful, etc., but not necessarily, but obvious tone of voi changes which are positive can be coded as such. Positives can be used in combination with management, cognitive, or both together.

- Negative or punishing comment--Usually this includes a command to stop something or shouting or speaking with raised voice; can be used alone when the behavior is not identified in the statement (see example) or with management, cognitive, or combination statements.

O The entire contact was student initiated (by a raised hand or formal student-teacher signal). This does not include eye contact or behavioral initiation such as misbehavior, but it does include arm tapping and skirt pulling, i.e., clear student initiation by physical signal or contact.

X For uncodeable or unheard statements, try to use only rarely.

✓ Teacher says nothing--and checks off work.

+MC A positive statement which has both management and cognitive overtones, such as "Great, you read all the pages assigned."

MC A neutral statement which deals with both management and cognitive behaviors, such as "You've begun reading; next, work with the cassettes."

-MC An overall negative interaction on either the management or the cognitive aspect of the interaction--theoretically a mixed contact, such as -M +C or +C -M, is possible, but I have never observed one.

In general, a contact is negative when the teacher tells a student to stop behaving in one way or another, neutral if behavior is requested, and positive if a behavior already done is praised. The decision as to whether or not the behavior is identified in the statement and, if so, whether it refers to cognitive or management behaviors.

7. On your observation sheet 6b, you will see four X's. The location of the X's is the location of the child you will interview. That is, if an X appears in the upper right hand corner of the third cell, find the child who is in the upper right hand corner of that cell in the class. Circle the sex of the child on the form and go and request to speak to him or her; if they refuse, record it as such and go to the next X.

153

You should interview two girls and two boys. That means, if your first two children are girls, then the next two should be boys closest to the X, etc. You should record the children's responses in list form--not complete sentences-- and only ask the additional questions if the child does not cover those areas by him/herself. (See below for example of child interview questions.)

Child Interview (2 boys and 2 girls randomly selected)

7 What do you do during PEP or IPI math ?

How do you get your assignment ?

How do you know what to do ?

Do you ever write your own prescription ?

What happens when you finish all of your work on your prescription sheet ?

APPENDIX D

SBOS Form and Definitions

used in the study,

The Use of Observational Data for Formative Evaluation of an Instructional Model

by

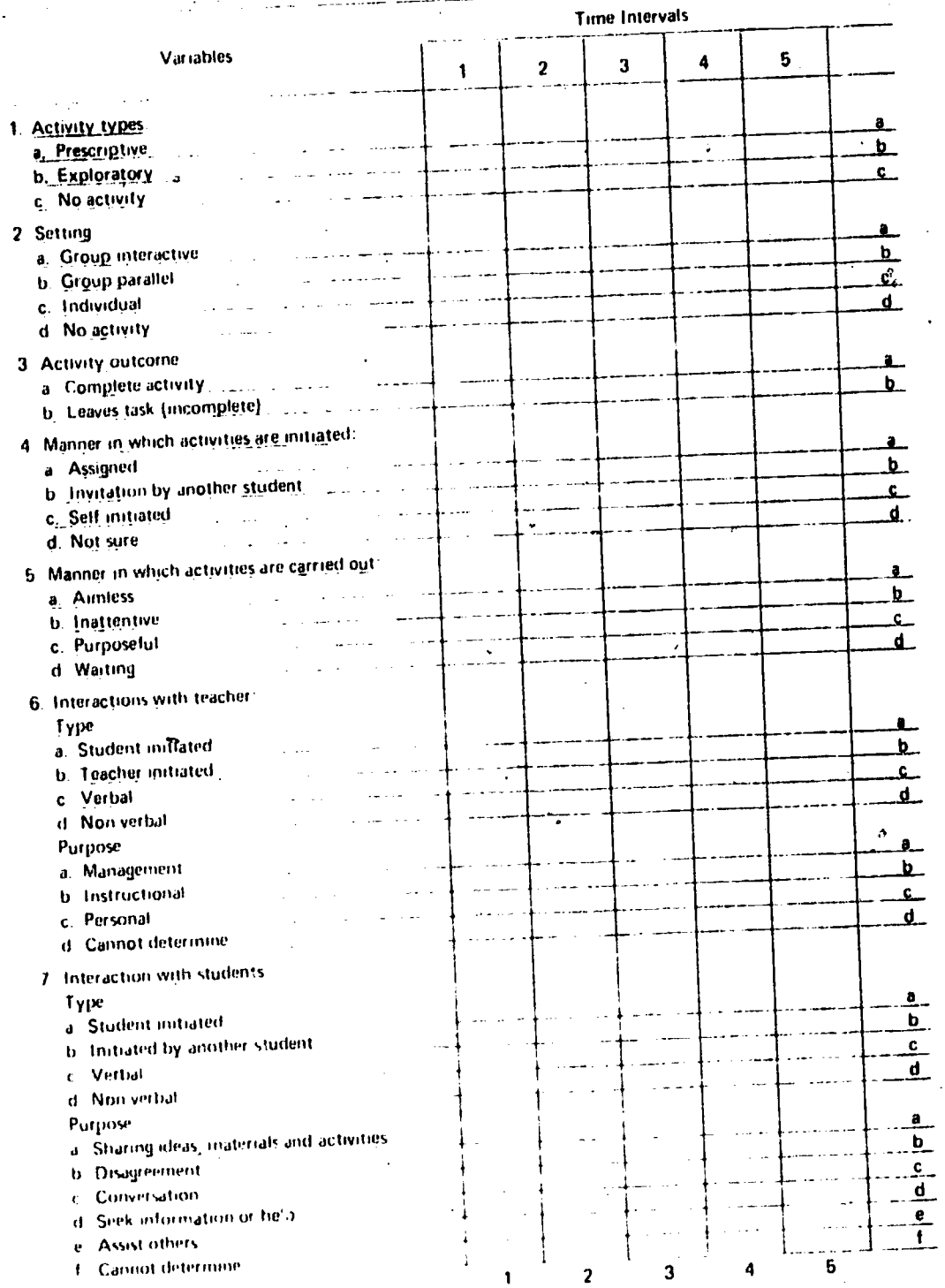
Margaret C. Wang

Learning Research and Development Center

University of Pittsburgh

APPENDIX D

Student Behavior Observation Schedule (SBOS)



Definition of Variables Included in the SBOS

1. Activity types

   a. Prescriptive--activities assigned by the teacher. These include tasks listed in students' prescription tickets, as well as on-the-spot assignments made by the teacher. Prescriptive activities are generally prescribed to teach specific skills, based on diagnostic tests of the LRDC individualized curricula.

   b. Exploratory--activities included in the exploratory component of the program. They generally include activities in creative arts, block construction activities, and extended activities that are related to such subject matters as science, math, reading, and prereading.

2. Setting

   a. Group interactive--engaging in an activity, among two or more students, sharing a common goal(s) and/or idea(s).

   b. Group parallel--engaging in an activity among two or more students, without sharing a common goal(s) and/or idea(s). A parallel group refers to location. For example, in block construction two students may be sharing blocks and sitting next to each other, but they are building different structures.

   c. Individual--student works alone.

   d. No activity--student is neither working on individual prescriptive, exploratory, nor group activity. Behavior such as clean up, waiting to be dismissed, going to lavatory, are considered under the "no activity" category.

3. Activity outcome

   a. Completed activity--(1) prescribed activity is recorded as completed when student actually puts the task he finished back on the shelf, and checked off by the teacher; (2) an exploratory activity

is recorded as completed when the student has completed the activity and when the teacher has checked off the activity on the exploratory ticket.

b. Leaves task (incomplete)--a student is recorded as having left the task incomplete when he leaves his prescriptive or exploratory, or group activity for more than an observation interval, or engages in a different activity before completing the task at hand.

4. Manner in which activities are initiated

a. Assigned--an activity is recorded as assigned to the student if the teacher directed him to do it.

b. Invitation by others--an activity is recorded as an invitation by others if the activity was suggested by another student or teache

c. Self-initiated--an activity is recorded as self-initiated if the student chose to do it on his own.

d. Not sure--an activity is recorded as "not sure" when the observ did not catch the manner in which the activity was initiated.

5. Manner in which activities are carried out

a. Aimless (wandering around without any activity)--a student's behavior is recorded as aimless if the student wanders around in the classroom not sure of what he wants to do or what he should be doing.

b. Unattentive (wonder, watching, while engaged in activity)--a student's behavior is recorded as unattentive if he stares into space, interacts with other students (verbal or nonverbal) on matters not related to the task at hand.

c. Purposeful--a student's behavior is recorded as purposeful if he is paying attention to what he should be doing.

d. Waiting--a student's behavior is recorded as waiting if he is sitting or standing with his hand raised for the purpose of getting the teacher's attention to check off his assignment, or ask questions, seek help, etc.

6. Interaction with teacher

a. Type:

(1) Student initiated--an interaction is recorded as student initiated if the student began the interaction either by speaking first or raising his hand.

(2) Teacher initiated--an interaction is recorded as teacher initiated if the teacher began the interaction.

(3) Verbal--an interaction is recorded as verbal if the student initiated or responded by speaking.

(4) Nonverbal--an interaction is recorded as nonverbal if the student responded or initiated physically such as nodding his head, laughing, or fighting without verbalizing.

b. Purpose:

(1) Management--an interaction is recorded as management if the purpose of the interaction dealt with noninstructional aspects, interactions not related to "academic" learning, such as asking and answering routine classroom management questions are recorded under this category.

(2) Instructional--an interaction is recorded as instructional if the purpose of the interaction is related to either prescriptive or exploratory activities. Interactions such as checking work, asking questions about student's learning task, tutoring, demonstrating, testing, and explaining are recorded under this category.

(3) Personal--an interaction is recorded as personal if it is related to personal matters (e.g., How do you feel today? Is your brother still in the hospital? What a pretty dress!).

7. Interactions with peers

a. Type:

(1) Student initiated--an interaction is recorded as student initiated if the student under observation began the interaction.

(2) Initiated by another student--an interaction is recorded as response to other student if the student under observation enters into an interaction which another student initiated.

(3) Verbal--an interaction is recorded as verbal if the student under observation responded or initiated by speaking.

(4) Nonverbal--an interaction is recorded as nonverbal if the student under observation responded or initiated physically such as by pushing other students, fighting, smiling, laughing, etc.

b. Purpose:

(1) Sharing--an interaction is recorded under sharing when the interaction involves sharing ideas, materials, activities, etc.

(2) Disagrees--an interaction is recorded as disagrees when the student under observation is either arguing or fighting with another student.

(3) Conversation--an interaction is recorded as conversation when the student under observation is talking to another student about personal matters not related to school work.

(4) Seek information--an interaction is recorded under seek information when the student asks other students to provide information on task related or nontask related matters.

(5) Assist others--an interaction is recorded as assisting others when the student under observation offers another student help or advice on management or instructional-learning matters.

APPENDIX E

Criteria for Observations Form

used in

A Developmental Study of Group Formation in Children

by

Donald M. Omark and Murray S. Edelman
University of Chicago

APPENDIX E

Criteria for Observations

		Dist.	Verb.	Agg.	Imit.	Int.	Comments
X	N1						
	N2						
Distance Covered	N3						
	Other						

Explanation

X	The child being observed
Distance Covered	This is an activity score and is an estimate of the total distance covered by the child during the 30 seconds he is observed.
N1, N2, N3	These are the three nearest neighbors, spatially, to X. They may or may not be interacting.
Distance	This is the distance which each is from X. This will be recorded as: (1/3m) C - actual contact with X. (1m) T - within touching distance if both N. and X raised their arms, but not touching (1-2m) S - normal speaking distance. (2m.) Y - yelling distance.
Verb	A check is put here if X or N is talking during the observation period. If direction of communication can be determined, an arrow is inserted in the box, e.g., means X talks to N = means both are talking to each other.
Agg	A check here means that an obvious physical aggressive encounter took place, e.g., hitting, punching, pulling down. If the fighting is one-sided, an arrow indicates the attacker
Imit	A check here indicates imitative behavior, e.g., two or more engaged in the same kind of action at the same time
Int	Physical interaction is occurring between X and N, but it is not aggressive or imitative. Games are included in this category. Describe under comments.
Comments	These lines are for comments about the gestures, imitative actions, etc. Interactions will be briefly described here. More detailed comments will permit later hypotheses generation.